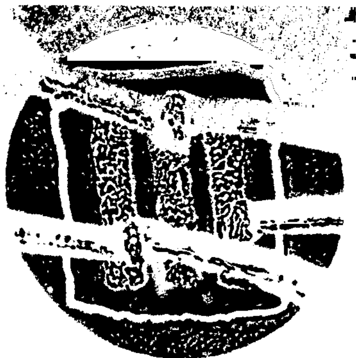
McGraw-Hill Encyclopedia

McGRAW-HILL BOOK COMPANY INC.

NEW YORK CHICAGO SAN FRANCISCO DALLAS TORONTO LONDON



of Science and Technology

AN INTERNATIONAL REFERENCE WORK

IN FIFTEEN VOLUMES INCLUDING AN INDEX

VOLUME 8 MAC-MYX



(LEFT) Microphotograph of new mesa transistor, which operates in the low microwave frequency region. The circle represents the cross-section diameter of an average human hair (*Bell Telephone Laboratories, Inc.*). (RIGHT) Surface of the moon.

Library of Congress Catalog Card Number: 60-11000

Suggestions to the Readers

The basic plan of the Encyclopedia is explained here in order to facilitate its use.

The subject matter of the various disciplines or branches of science and technology is organized systematically: a general article provides a broad survey of the field, and a number of separate articles, alphabetically arranged, cover its main subdivisions and more specific aspects.

Cross references guide the reader from the general articles to the other articles into which the subject is subdivided, and from these to articles on more highly specialized phases of the subject. The cross references—there are about 50,000 of them—are printed in small capital letters so that they can be easily recognized. By means of the cross references a reader may find his way from ELECTRICAL ENGINEERING, through ELECTRONICS and VACUUM TUBE, to ELECTRON MOTION IN VACUUM or ELECTRON EMISSION. Or, following another line of cross references, the reader would be led to ELECTRIC POWER SYSTEMS, TRANSMISSION LINES, ELECTROMAGNETIC WAVE, and so on.

In general, each article begins with a definition of the title that states its scope and coverage. Usually, only the scientific or technological sense is discussed. Most of the articles, after this statement, go on to increasingly complex and detailed considerations. A reader thus needs to proceed only as far as his inclinations and requirements dictate.

The Index, Volume 15, should be consulted to locate the discussion of topics covered in the Encyclopedia but not given in separate entries.

Every phylum, class, and order in the plant and animal kingdoms is allotted a separate article. Many of the more common families, genera, and species are covered either in one of the order articles or in a separate article under its own scientific or common name.

The adjectives electric and electrical are used in the following senses. Electric—containing, producing, arising from, actuated by, or carrying electricity, or capable of doing so; as, for instance, electric generator, electric motor, electric wiring. Electrical—related to, pertaining to, or associated with electricity, but not having its properties or characteristics; as, for example, electrical code, electrical engineering.

Words used as titles are, wherever possible, given in the singular to permit a consistent alphabetic arrangement. Titles are alphabetized by word and not by letter, for example,

Earth sciences
Earth tides
Earthmover
Earthquake

A word used as a noun precedes the same word used adjectivally, thus:

Mercury (element)
Mercury (planet)
Mercury battery

or

Circuit, electronic
Circuit breaker

Hyphened terms are alphabetized as single words; for example,

Animal virus
Animal-feed composition

Most of the longer articles contain bibliographies citing useful sources of further information. For additional bibliographical citations, the reader should refer to related articles (as indicated by the cross references in the article). Bibliographies are placed at the ends of articles or sometimes at the ends of major sections in long articles.

A list of initials and names of the contributors to the Encyclopedia is to be found in Volume 15. This list will permit quick identification of a contributor's initials after an article. Immediately following this list is a second list of encyclopedia contributors with their affiliations and the titles of articles each has written for the Encyclopedia.

McGraw-Hill Encyclopedia of Science and Technology



Mach number

In fluid mechanics, the ratio v/c of the free stream velocity v to the velocity of sound c in the fluid at the same condition, such as temperature and pressure. Mach number is also the ratio of the inertia force of the fluid to the force of compressibility or the elastic force. In most fluid systems, compressibility effects become important for values of the Mach number greater than about 0.3. A body moving through a fluid at a velocity less than sonic is preceded by a region of gradually varying density and pressure that controls the flow around the body. At Mach numbers equal to or greater than unity the gradual transition of pressure cannot exist and shock waves, or regions of abruptly altered pressure and density, form at critical sections on or near the surface of the body and extend outward. As a result, the fluid force pattern on the moving body is markedly different at supersonic velocities from the pattern at subsonic velocities (*see* STREAMLINING). The theory for calculating the pressure patterns is well established for ideal fluids (*see* SHOCK WAVE).

When compressibility effects alone are significant, geometrically similar bodies will develop identical flow and shock wave patterns when operated at equal Mach numbers (*see* DYNAMIC SIMILARITY; MODEL THEORY). However, in the hypersonic region, generally considered to mean a Mach number of 5 or more, there is appreciable interaction between the boundary layer and the shock pattern. When this occurs, the Reynolds number as well as the Mach number is significant and the similitude requirements become more difficult to satisfy.

Techniques of transformation of variables have been devised whereby boundary conditions may be altered to accommodate nonconformance to the Mach requirement. That is, data from incompressible flow may be used to give accurate predictions under compressible-flow conditions using distortion of geometry to balance distortion of velocity.

[G.M.]

Machinability (metals)

The capacity of a metal to respond to machining is not a basic standard but is relative, and the basis for any rating should be defined clearly. Good machinability involves several criteria: long tool life for a given cutting speed, high surface quality of the machined surface, well-broken-up chips for easy disposal, low power consumption in removing a given quantity of metal per tool grind, and the removal of each unit of metal at the lowest cost.

Finished metal parts are usually machined to final size and shape from stock previously rough-formed by casting, forging, rolling, or extrusion. The machining is done with tools having sharp cutting edges made of materials harder than the metal to be cut. There are a variety of machining operations, such as turning, boring, facing, shaping, and planing with single-point tools (Fig. 1); milling flat or formed surfaces with multipoint milling cutters; originating holes with two-lipped drills; enlarging holes to size with reamers; broaching internal or external surfaces with multitooth broach cutters; and threading, sawing, and grinding, with appropriate tools. More than $10,000,000,000 worth of machining work is done annually in the United States.

In every machining operation several factors are highly important: the material to be cut, the tool to do the cutting, the fluid to facilitate cutting, the machine tool on which the work is done, and the process.

The rated machinability of two or more metals (or tools or fluids) may vary for different cutting processes, such as heavy or light turning, form turning, milling, drilling, sawing, broaching, or grinding. Almost always one or more objectives must be sacrificed to obtain others, but in commercial production, economy is paramount.

There is no single rapid test to determine the machinability of a metal, tool, or cutting fluid for commercial use. One type of metal out of several may give the best tool life, but a second may provide better surface quality, and a third may yield

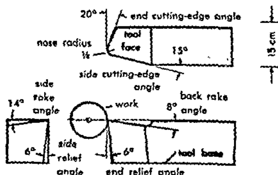


Fig 1. The nomenclature and designation of a typical solid tool of high-speed steel ground for turning steel of 250–300 Brinell.

better broken up chips or require less power. By keeping two of the elements, such as tool and cutting fluid, constant and varying the third, metal cut, an optimum combination can be found. Ratings are based on several factors: the relation of cutting speed to tool life for conditions simulating practice; cutting forces, gross power, or unit net power (hp per cu in. per min); cutting time under constant feeding force in turning, drilling, or sawing; surface quality, cutting temperatures at the tool point; heat-power balance of chips removed; torque and thrust in drilling and tapping, and hardness, or strength, of the metal being machined. Graphs of test results are shown in Figs. 2 and 3.

Materials. Hardness, tensile strength, and ductility indicate only approximate machinability ratings. In general, harder materials require lower cutting speeds for a given tool life and greater power, but give better surface finish on the work. Metals having the same hardness but varying microstructures will cause significant differences in tool wear and life. Similar microstructures, regardless of chemical analyses, generally cause equivalent tool wear.

Brittle metals require a tool of high hardness and strength, and low rake, so as to present a blunt cutting edge (see Fig. 1). The chips are removed by pressures on the tool face just back of the cutting edge, in small, broken, but only slightly distorted particles. The tool scrapes over the surface to even it up, causing tool failure by flank abrasion. In ductile metals the chips are removed as continuous ribbons and with a great deal of plastic distortion, producing excessive heat in the shear zone ahead of the cutting edge and still more as the chip slides over the tool face. This heat causes the tool to fail by flank wear and cupping. The chip is work-hardened and becomes harder than the original metal. Stainless steels work-harden excessively, whereas titanium does not.

For mass production, steels are given one heat-treatment to provide a structure most suitable for rough machining, a second for finish machining, or form machining, and a third to assure the hardness, strength, or ductility required for the end product. Steels may be fully annealed to give a lamellar pearlitic structure of low hardness; normalized to give a fine, tempered, martensitic structure; specially annealed to give a spheroidized structure (best for machining high-carbon steels); or quenched and tempered to give a martensitic structure for high hardness and strength.

Soft metals are ductile and are readily machinable but usually give a poor surface finish. Their machinability is improved by cold drawing to reduce ductility. Metals of very high hardness are difficult and expensive to machine even with low speeds. Grinding, machining at high temperatures, electro-discharge machining, and ultrasonic machining are often used with hard metals. These processes are finding increasing application in industries working the high-temperature and high-strength metals. Barrel tumbling with a variety of abrasives and chemicals is also used to deburr parts and produce a high-quality surface finish while maintaining dimensional accuracy. Some steels that contain manganese sulfide, and that are used for quantity production of pieces that do not require high thermal or mechanical properties, can be machined at higher cutting speeds. Lead is sometimes added to steel and brass to make them free-machining. Most special free-cutting steels, brass, and aluminum are machined in the cold-drawn state in high-speed automatic screw machines.

Cutting tools. These are made of a variety of materials to meet the many requirements of the metal-cutting industries. Carbon steel tools having low initial cost are used to machine the light metals, or free-cutting steels, with light cuts at low speeds. They retain their hardness up to temperatures of 400°F. High-speed steel tools of many types are the work horses of industry. Their cost is several times that of carbon steel tools, but they permit cutting speeds about twice those of carbon steel and withstand temperatures up to 1000°F. *Tungsten-base* and molybdenum-base high-speed steels, each having chromium and vanadium and sometimes cobalt and other metals added, fill special requirements.

Tools of sintered carbide are still more expensive than those of high-speed steel. They have greater hardness with working-temperature limits up to

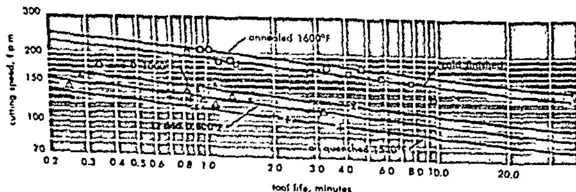


Fig. 2 Relation of tool life to cutting speed when turning annealed cold drawn and heat treated AISI 1045 steel, dry

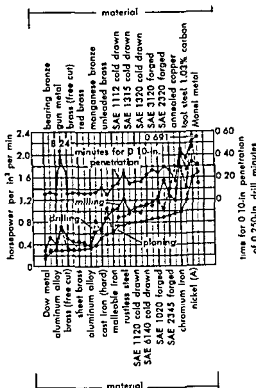


Fig., 3. Net values of horsepower per cubic inch of metal cut per minute at the cutter for a variety of metals as determined for drilling, milling, and planing with metal-cutting dynamometers. The drill used was a ¾-in. diameter, 30° helix, and was operated at 153 rev/min and 0.012 in./rev feed. The planing tool was of the end-cutting type ½ in. wide, having 15° back rake, no side rake, and operated at a speed of 20 ft/min when taking a depth of cut per stroke of 0.010 in. The milling cutter was of the end-cutting type 0.25 in. wide, 3.5 in in diameter, with 15° back rake, no side rake when taking a depth of cut of 0.125 in., and a feed of 0.010 in. per tooth. The penetrator drill indicating machinability was a ¼-in. diameter, 24° helix and operated under a feed load of 94 lb at 500 rev/min.

1800°F. Tools of these metals will machine steels of average hardness 4–8 times as fast as those of high-speed steel. They will machine very hard metals satisfactorily. Being less ductile than the tool steels, they must be carefully designed and supported to prevent chipping or fracture. Because of their high cost they are used as tips, small formed pieces which are brazed or clamped onto rigid steel shanks. Present practice in high production is to use small, thin, indexible tips of triangular, square, or circular profile, clamped to strong shanks. When the cutting points on both faces are worn, the tip is thrown away rather than resharpened. This reduces or eliminates the expensive diamond-wheel grinding.

Industrial diamonds are also used to machine metals at light cuts and high speeds up to 1000 ft/min and to produce high dimensional accuracy and superior surface finish. Ceramic materials, called cermets, and cemented oxides permit high cutting speeds (1000–15,000 ft/min) with light cuts. Speeds of 1000–3000 ft/min are more practical for the thin, indexing, throw-away types.

Cutting fluids. Lubricants or coolants are applied to the work and tool to assist in cutting operations. They include air, used as suction or blast; water containing an alkali; an emulsion of soluble oil and water; straight mineral oils or mixtures of mineral oils with fatty oils; and straight oils or mixed oils that have been sulfurized, chlorinated or both. Water is the best coolant, but emulsions that provide some lubrication and anticorrosive properties are generally used for all high-speed machining. They increase cutting speeds by 10 to 35%. Low-viscosity oils are used in complicated machines to lubricate the machine and cool the tool particularly in machining brass, aluminum, and magnesium. Water aggravates the fire hazard of magnesium. The sulfurized oils are used for low cutting speeds on operations such as threading, gear cutting, reaming, and broaching primarily to provide good surface finish and high dimensional accuracy. Oils increase cutting speeds by 8 to 15%. All fluids must be applied copiously to the tool. The precooling of the fluids, particularly the emulsions, increases tool life.

Since 1935 many studies have been made to establish a mathematical relationship between cutting force, chip compression, coefficient of friction of the chip sliding over the tool face, and tool cutting angles. Many useful conclusions have been reached, but, as in so many fields, there is still a gap between theory and practice. *See* MACHINING OPERATIONS; METAL, MECHANICAL PROPERTIES OF; PLASTIC DEFORMATION OF METAL. [O.W.B.]

Bibliography: American Society of Mechanical Engineers, *Manual on the Cutting of Metals*, 1952; American Society for Metals, *Metals Handbook*, 1948 and supplements 1954, 1955; American Society of Tool Engineers, *Tool Engineers' Handbook*, 1949; O. W. Boston, *Metal Processing*, 2d ed., 1951.

Machine

A combination of rigid or resistant bodies having definite motions and capable of performing useful work. The term mechanism is closely related but applies only to the physical arrangement that provides for the definite motions of the parts of a machine. For example, a wrist watch is a mechanism, but it does no useful work and thus is not a machine.

Machines vary widely in appearance, function, and complexity from the simple hand-operated paper punch to the ocean liner, which is itself composed of many simple and complex machines. No matter how complicated in appearance, every machine may be broken down into smaller and smaller assemblies until an analysis of the operation becomes dependent upon an understanding of a few basic concepts, most of which [illegible] elementary physics. *See* SIMPLE [illegible]

Machine design

Application of science and invention to the development and construction of machines. An understanding of the basic laws of nature is essential to a proper perspective in the approach to machine design. Knowledge of the past development of machine elements makes possible their effective application. Inventiveness consists of producing new combinations of old elements or, where extreme need arises, of exercising genius either in breaking the bounds of convention, or in evolving new principles not hitherto applied or known.

In machine design, accomplishment takes on two forms: one is the drawings and blueprints, which completely describe the machine, and the other is the assembled product. In addition, most machines go through periods of evolution, and later models may show little outward similarity to the original design.

Machine design consists of the conception of a machine that will meet a specific need. Before constructing a machine to fulfill the need, the designer must thoroughly understand the application, and mentally modify an old machine or devise a new machine as required. He estimates a certain cost for the machine, and a probable time for its construction. He envisions the materials required, the equipment necessary for its manufacture and testing, and the final operation in meeting the original need. If the machine is desirable, he converts his thoughts into drawings and materials, and follows through to its fabrication. In time, the machine may become obsolete due to advances in the technology; it may then be rebuilt or replaced, possibly under the direction of the original designer.

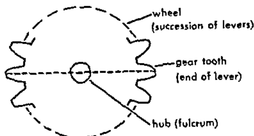
The working tools in machine design are an understanding of the basic elements of machines that have been developed in the past and a thorough knowledge of the mechanical fields of science including mathematics, physics, statics and dynamics, strength of materials, kinematics, mechanisms, and the laboratories associated with them. See MECHANISMS. [J.J.R.]

Bibliography: S. J. Berard, E. O. Waters, and C. W. Phelps, *Principles of Machine Design*, 1955.

Machine elements

Elementary mechanical parts used as building blocks for the construction of most devices, apparatus and machinery. The gradual development of these building blocks, following the invention of the roller or wheel and the arm or lever in ancient times, brought about the Industrial Revolution, starting with the assembly of James Watt's engine for harnessing the force of steam and proceeding into the advanced mechanization of present automatic control.

The most common example of a machine element is a gear, which, fundamentally, is a combination of the wheel and the lever to form a toothed wheel as illustrated. The rotation of this gear on a hub or



Gear is a machine element combining features of the wheel and the lever.

shaft drives other gears which may rotate faster or slower, depending upon the number of teeth on the basic wheels. The material from which the gear is made establishes its strength, and the hardness of its surface determines its resistance to wear. Knowledge of the forces on the gear makes possible the determination of its size. Changes in its shape allow modifications in its use. These applications, as in most machine elements, have developed into many standard forms, such as spur, bevel, helical, and worm gears (*see* GEAR). Each of these forms has required the development of a special technology for its production and use.

Other fundamental machine elements have evolved from wheels and levers. A wheel must have a shaft on which it may rotate. The wheel is fastened to the shaft with a key, and the shaft is joined to other shafts with couplings. The shaft must rest in bearings such as journal bearings, ball bearings, or roller bearings. The shaft may be started by a clutch or stopped with a brake. It may be turned by a pulley with a flat belt, a V belt, or a chain or a rope connecting it to a pulley on a second shaft. The supporting structure may be assembled with bolts or rivets or by welding. Proper application of these machine elements depends upon a knowledge of the forces on the structure and the strength of the materials employed. In the design, calculations must accommodate the forces to the materials in the simplest construction.

Other machine elements have been evolved whose applications are more specific in construction. *See* BEARING, ANTIFRICTION; BRAKE; CAM MECHANISM; CLUTCH; COUPLING; FLYWHEEL; FOUR-BAR LINKAGE; GOVERNOR; SCREW FASTENER; SHAFTING; SPRING (MECHANICAL).

Machine parts which are commonly used have been developed into standardized designs. Manufacturing specialists have concentrated upon the development of standard elements and have mass-produced these parts with a high degree of perfection at reduced cost. Standard elements, as applied in machine design, may be modified as desired, although certain ones, through the hazards to safety under improper use, must be modified only in line with the requirements of codes established by regulating bodies. *See* DESIGN STANDARDS. [J.J.R.]

Bibliography: V. M. Faires, *Design of Machine Elements*, 3d ed., 1955.

Machine key

The most common function of a key is to prevent relative rotation of a shaft and the member to which it is connected, such as the hub of a gear, pulley, or crank. Many types of keys are available and the choice in any installation depends on such factors as power requirements, tightness of fit, stability of connection, and cost. For light power requirements a setscrew may be tightened against the round shaft or against a flat spot on the shaft. For most requirements a positive connection, such as by a key, is necessary. A setscrew is frequently used to seat the key and to prevent axial motion.

Square keys are common in general industrial machinery (Fig. 1*a*). Flat keys are used where added stability of the connection is desired, as in machine tools (Fig. 1*b*). Square or flat keys may be of uniform cross section or they may be tapered. In tapered keys the width is uniform and the height of the key tapers. Tapered keys may have gib heads to facilitate removal (Fig. 1*c*).

The Kennedy key is used for heavy duty and consists of two keys driven 90° apart (Fig. 1*d*). The hub is bored to fit the shaft and is then rebored slightly off center. The keys force the shaft and hub into concentric position.

The Woodruff key requires a key seat formed by a special side-milling cutter (Fig. 1*e*). This key will align itself in the key seat. It has the disadvantage of weakening the shaft more than a straight key.

The round key, or pin, introduces less stress concentration at the key seat in the shaft and is satisfactory except for the necessity of drilling the hole to accommodate the pin after assembly of the hub and shaft (Fig. 1*f*). This may be a disadvantage in production and prevents interchangeability.

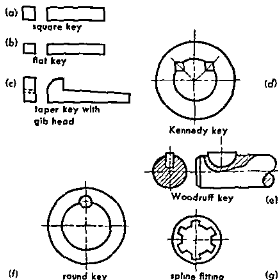


Fig. 1. (a-g) Types of keys. (*From P. H Black, Machine Design, 2d ed., McGraw-Hill, 1955*)

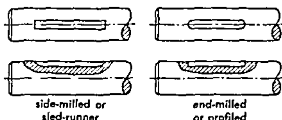


Fig. 2. Types of keyways (*From P. H. Black, Machine Design, 2d ed., McGraw-Hill, 1955*)

A spline fitting is composed of a splined shaft

shaft. Splined fittings are adaptable to mass production and are used where radial space must be conserved.

Straight-side splines are being replaced at an increasing rate by stub involute splines. These splines have the advantages of greater strength, a self-centering feature, and production economy.

Keyways for straight keys are formed either by a side-milling cutter which forms a sled-runner keyway, or by an end-milling cutter which forms profiled keyways (Fig. 2). The sled-runner keyway requires a longer space between the end of the key and the end of the keyway than does an end-milled keyway. This favors the end milled keyway in locations near a shoulder. However, the end milled keyway reduces the endurance limit of a shaft more than does the sled-runner type.

Feather keys are used where it is necessary to slide a keyed gear or pulley along the shaft. The key is generally tight in the shaft and with clearance between key and hub keyway. When a gear or pulley must be moved axially along the shaft while power is being transmitted, it is desirable to provide for a minimum of force necessary to move the hub along the shaft. The use of two feather keys equally spaced reduces the necessary axial force to half that for one key. [P.H.B.]

Bibliography: American Standards Association, *Involute Splines, Side Bearing*, B5.15–1950; ASA, *Woodruff Keys, Keyslots and Cutters*, B17f–1930; P. H. Black, *Machine Design*, 2d ed., 1955.

Machinery

A group of parts arranged to perform a useful function. Normally some of the parts are capable of motion; others are stationary and provide a frame for the moving parts. The terms machine and machinery are so closely related as to be almost synonymous; however, machinery has a plural implication, suggesting more than one machine. Common examples of machinery include automobiles, clothes washers, and airplanes; machinery differs greatly in number of parts and complexity.

Some machinery simply provides a mechanical advantage for human effort. Other machinery performs functions that no human being can do for

long-sustained periods. A jackscrew does nothing until a man pulls on a lever; then he is able to move objects many times his weight. Conversely, an internal combustion engine can run unattended for hours, requiring only the press of a button to start it.

The need for machinery usually stems from a desire to do a job at less cost. Evolution of machinery for a certain function may be gradual or rapid. If only small quantities of a product are needed, it is likely that machinery used in making the product will not change rapidly or possess the highest degree of automation. On the other hand, machinery used in making automobiles has evolved into some of the most complex automatic machines in existence. See MACHINE DESIGN; MECHANICAL ENGINEERING [R.S.SH.]

Machining operations

Methods for the removal of material from a workpiece by the use of a cutting medium. Machining operations are usually carried out in a power-driven arrangement which gives the piece a desired shape or finish.

Cutting tools. Single- or multiple-point cutting tools, including broaching tools, milling cutters, honing stones, abrasives, and saw blades, are common forms of cutting devices. Although varied in configuration, the basic cutting action performed by each is much the same. Some tools remove material in large chips, shavings, or large pieces; oth-[illegible]

in some cases the workpieces, and furnish the power for cutting, vary greatly in size and configuration. Machines vary from small hand-held drilling or grinding devices to large, automatic, multioperation machine tools. Machining operations commonly performed are turning and facing, boring, milling, sawing, broaching, shaping, planing, drilling, threading, tapping, the various types of gear cutting, grinding, honing, lapping, superfinishing, buffing, polishing, and nibbling. Some of these overlap each other; operations such as polishing may not be commonly thought of as *machining*.

Motion between tool and work. In power machining operations there must be a relative movement between cutting tool and workpiece. During some machining operations the workpiece moves; in others only the cutting tool moves, frequently both tool and workpiece are in motion. The various machining operations that employ cutting tools use either a rotating or a traversing motion. For example, if either a turning or boring operation is performed on a lathe, the workpiece rotates while the cutting tool traverses longitudinally along or through the piece. On a milling machine the opposite may be true; the cutting tool, or milling cutter, rotates while the workpiece is clamped to a table and made to traverse the path of the cutter teeth.

In certain cases an operation is accomplished by motion of the tool on one machine and by motion of the workpiece on another machine. An example is a drilling operation performed on a turret lathe; the drill is traversed without rotating, while the workpiece revolves. Had the operation been done on a drill press, the workpiece would have remained stationary while the drill rotated as it advanced into the work.

Basically in any machining operation in which a cutting tool is used, one or more relative motions between the workpiece and the cutting device are necessary.

Many particular machining operations can be performed on a workpiece most easily or most economically by a certain machine tool, although it is frequently possible to perform the operations on more than one type of equipment. Of primary consideration in fitting the job to the machine is the nature of the required operation. Surfaces to be machined may be considered as being fl[illegible] tou[illegible] terr[illegible] to these, there are the surfaces or configurations produced by specialized operations such as threading, tapping, and hobbing.

Choice of machine method. Once the machining requirements have been assessed, the job is fitted to the most appropriate machine available. The basic rotary and traverse machining motions provide a primary step in relating the work requirements to the machine. The accompanying table relates these motions to many of the common machine tools and also to the various surface configurations that the particular machine is capable of producing.

Machines for cylindrical surfaces. The table shows that machines 1–7 are capable of machining or finishing some type of cylindrical surface and most of them can machine other types of surface as well. Several of these first seven machines are extremely versatile. The horizontal boring mill can, in addition to turning and boring, perform milling operations equally well.

The difference in the work produced by these machines lies mainly in (1) the size of either the bore or the outside diameter of the piece, (2) the physical size of the workpiece handled, and (3) the tool used. The drilling machine, for example, using a twist drill for a tool, is able at best to produce holes only a few inches in diameter. While radial drilling machines are built to handle pieces of considerable size, nevertheless machine drilling usually refers to pieces of small to medium size. By contrast, large vertical boring mills, using single-point [illegible] ...nishes an existing one; yet both machine internally cylindrical surfaces or bores.

Machine 7, the cylindrical grinder, is designed to refine or finish externally cylindrical surfaces.

Basic machine tool functions

No.	Machine type	Tool movement	Work movement	Machined surfaces
1	Drilling machine	Rotate and traverse	Stationary	Internally cylindrical; specialized
2	Horizontal lathe (engine, turret, screw machine)	Traverse (longitudinally and radially)	Rotate	Externally cylindrical; internally cylindrical; shoulders or ends of cylinders; specialized surfaces
3	Vertical lathe	Traverse (longitudinally and radially)	Rotate	Same as horizontal lathe
4	Boring machine (horizontal)	Rotate and traverse	Traverse	Externally cylindrical; internally cylindrical; shoulders or ends of cylinders; flat and flat contoured; specialized surfaces
5	Boring machine (vertical)	Traverse	Rotate	Externally cylindrical; internally cylindrical; shoulders or ends of cylinders
6	Broaching machines (horizontal and vertical)	Traverse and stationary	Traverse and stationary	Flat and flat contoured; curved; cylindrical
7	*Grinder (cylindrical)*	*Rotate and traverse*	*Rotate (also traverse on centerless type)*	*Externally cylindrical*
8	Milling machine	Rotate	Traverse	Flat and flat contoured; curve contoured
9	Planer (vertical)	Traverse	Traverse	Same as milling machine
10	Shaper	Traverse	Stationary	Same as milling machine
11	Planer (horizontal)	Stationary (planer miller rotates)	Traverse	Same as milling machine
12	Grinder (surface)	Rotate	Traverse	Flat surfaces
13	Nibbling machine	Traverse	Traverse	Flat or curved
14	Saw (circular)	Rotate	Traverse	Flat
15	Saw (band)	Traverse	Traverse	Flat or curved
16	Flame cutter	Traverse	Stationary	Flat or curved

Machines for flat surfaces. Machines 8–12 on the chart are all capable of machining flat or flat contoured surfaces. These machines are usually able to produce curved contoured surfaces by means of contour cutters, two-dimensional movement of either the tool or workpiece, or by special fixtures.

The machines vary in size, capacity, and type of tool. Common milling machines, with rotating cutters, are used primarily for smaller items. Horizontal planers and milling planers, with single-point and rotating face mills respectively, are built in models designed to handle pieces of all sizes. Versatile machines such as boring mills, broaching machines, and vertical turret lathes can finish flat surfaces. Machine 11, the surface grinder, as in the case of the cylindrical grinder, actually refines or finishes flat surfaces.

Machines for parting or trimming. Machines 13–16 are used for parting stock or removing excess material. The nibbling machine with its punching action is able to cut out irregular shapes from relatively thin sheets of material, while a tool such as the flame cutter can cut through several inches of steel plate. The power saws fall between these two extremes.

The factors that differentiate one machine from another are size, capabilities, workpiece requirements, machine speed, rigidity, and work loads. These and other factors enter into scheduling work on one type of machine in preference to another. *See* PRODUCTION METHODS.

Single-point tool. Basic among the various tool forms used in power machining is the single-point tool (Fig. 1). A form of this tool is used on lathes.

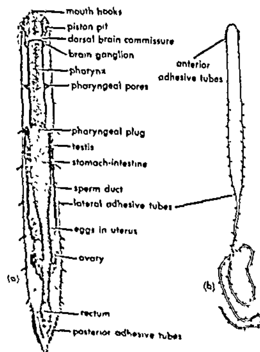


Fig. 1. Inspection of a single-point cutting tool on an optical comparator and measuring machine. (Jones and Lamson Machine Co.)

Condensing vapors cannot be measured with a McLeod gage, although an absorber or cold trap may permit measurement of partially condensable mixtures. [B.D.H.; H.C.P.]

Macrodasyoidea

An order of the class Gastrotricha in which the adhesive tubes are distinctive. There are as many as 250 of these structures. They are cylindrical, cuticular projections which are moved by delicate muscle strands. Minute glands, clustered ventrally on the head surface or disposed in lateral or transverse longitudinal rows, produce a sticky secretion. They are part of the syncytial epidermis lying beneath them. Head differentiation is not usual. The sides of the body are parallel, and the shape of the animal is elongate. A few species show anterior



Macrodasyoidea. (a) *Macrodasys*, marine gastrotrich, with lateral adhesive tubes. (b) *Urodasys*, marine gastrotrich, with long tail. (After A. Remane, 1926, from L. H. Hyman, *The Invertebrates*, vol. 3, McGraw-Hill, 1951)

red pigment spots, possibly light sensitive. Protonephridia for elimination of metabolic waste are lacking. Hermaphroditism occurs in members of this group. *Macrodasys*, *Turbanella*, and *Thaumastoderma* are representative genera, comprising numerous Old World species. A few are now being reported from the United States with unpublished data known to exist. See GASTROTRICHA. [C.E.P.]

Macroevolution

The larger course of evolution by which the categories of animal and plant classification above the species have been evolved from each other and have differentiated into the forms within each. Only that part of evolution for which concrete evidence exists will be considered.

The organisms observed, alive or as fossils, can be classified into large groups or phyla. Within each of these the body is organized in variations of a single fundamental type. Such groups are the annelids, mollusks, echinoderms, and other phyla. Macroevolution is the process by which evolution has taken place within these phyla. In some cases evidence can be given that two or more such phyla have earlier evolved from a single ancestral group.

The evidence for the reality of macroevolution is derived from many sources. Paleontology provides evidence of the course of evolution from fossils. Fossils preserved at intervals of geological time may be compared and deductions made of their relationships. Thus, phylogenetic trees may be constructed. Studies of comparative morphology and embryology give evidence of many kinds from the similarities and dissimilarities of living animals. Geographical distribution of forms, together with the known history of the lands, gives data which should be in agreement with beliefs concerning the evolution of the organisms.

Speciation or microevolution has continued throughout evolution and provides its background (see SPECIATION). But the question remains whether the course of change in the long progress of macroevolution has been controlled in other ways besides the mechanisms considered in speciation. Another question for discussion is whether macroevolution is merely long-continued speciation. Before the evidence on this question can be considered, some characters of the general nature of macroevolution must be noted.

GENERAL CONSIDERATIONS

Throughout the course of evolution, change has occurred during the life of organisms. These organisms must always be competent to survive in competition with other organisms. The whole biology evolves; evolution is a phenomenon in natural history. Also, change of habit and habitat normally accompanies the evolution of structure. It is probably without meaning to ask which of these precedes the other.

Types of change. In all evolution one may distinguish two types of change, cladogenesis and phyletic evolution. Cladogenesis is the type of evolution associated with altered habit and habitat, usually in parts of a species separated from the rest. Adaptation to the new conditions results in division of the originally homogeneous species into distinguishable parts. Phyletic evolution is the gradual evolution of a population in its environment without division into isolated parts. This consists partly in improved adaptation to the conditions, partly in readaptation when the conditions change, as they frequently do in natural environments. Usually change in environmental conditions is gradual, but sudden and irreversible change sometimes occurs ... isms ... into new

Basic machine tool functions

No.	Machine type	Tool movement	Work movement	Machined surfaces
1	Drilling machine	Rotate and traverse	Stationary	Internally cylindrical; specialized
2	Horizontal lathe (engine, turret, screw machine)	Traverse (longitudinally and radially)	Rotate	Externally cylindrical; internally cylindrical; shoulders or ends of cylinders; specialized surfaces
3	Vertical lathe	Traverse (longitudinally and radially)	Rotate	Same as horizontal lathe
4	Boring machine (horizontal)	Rotate and traverse	Traverse	Externally cylindrical; internally cylindrical; shoulders or ends of cylinders; flat and flat contoured; specialized surfaces
5	Boring machine (vertical)	Traverse	Rotate	Externally cylindrical; internally cylindrical; shoulders or ends of cylinders
6	Broaching machines (horizontal and vertical)	Traverse and stationary	Traverse and stationary	Flat and flat contoured; curved; cylindrical
7	Grinder (cylindrical)	Rotate and traverse	Rotate (also traverse on centerless type)	Externally cylindrical
8	Milling machine	Rotate	Traverse	Flat and flat contoured; curve contoured
9	Planer (vertical)	Traverse	Traverse	Same as milling machine
10	Shaper	Traverse	Stationary	Same as milling machine
11	Planer (horizontal)	Stationary (planer miller rotates)	Traverse	Same as milling machine
12	Grinder (surface)	Rotate	Traverse	Flat surfaces
13	Nibbling machine	Traverse	Traverse	Flat or curved
14	Saw (circular)	Rotate	Traverse	Flat
15	Saw (band)	Traverse	Traverse	Flat or curved
16	Flame cutter	Traverse	Stationary	Flat or curved

Machines for flat surfaces. Machines 8–12 on the chart are all capable of machining flat or flat contoured surfaces. These machines are usually able to produce curved contoured surfaces by means of contour cutters, two-dimensional movement of either the tool or workpiece, or by special fixtures.

The machines vary in size, capacity, and type of tool. Common milling machines, with rotating cutters, are used primarily for smaller items. Horizontal planers and milling planers, with single-point and rotating face mills respectively, are built in models designed to handle pieces of all sizes. Versatile machines such as boring mills, broaching machines, and vertical turret lathes can finish flat surfaces. Machine 11, the surface grinder, as in the case of the cylindrical grinder, actually refines or finishes flat surfaces.

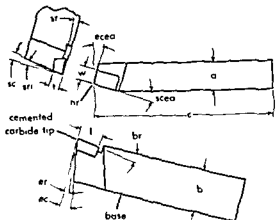
Machines for parting or trimming. Machines 13–16 are used for parting stock or removing excess material. The nibbling machine with its punching action is able to cut out irregular shapes from relatively thin sheets of material, while a tool such as the flame cutter can cut through several inches of steel plate. The power saws fall between these two extremes.

The factors that differentiate one machine from another are size, capabilities, workpiece requirements, machine speed, rigidity, and work loads. These and other factors enter into scheduling work on one type of machine in preference to another. See PRODUCTION METHODS.

Single-point tool. Basic among the various tool forms used in power machining is the single-point tool (Fig. 1). A form of this tool is used on lathes,



Fig. 1. Inspection of a single-point cutting tool on an optical comparator and measuring machine. (Jones and Lamson Machine Co.)



ecea (end cutting edge angle)
scea (side cutting edge angle)
sri (side relief)
sc (side clearance)
er (end relief)
ec (end clearance)
br (back rake)
sr (side rake)
a (shank width)
b (shank height)
c (shank length)
t (tip thickness)
w (tip width)
l (tip length)
nr (nose radius)

Fig. 2 Single-point tool nomenclature diagram. (Kennametal Inc.)

boring mills, shapers, and planers, with variations employed on other machines. Single point tools vary in design, each adapted to the operation for which it is intended. The tool may be straight, bent, offset, or have some other special form that will enable its cutting tip to reach the desired area of machining (Fig. 2). The tool may be made from one solid piece of tool steel with the cutting edge ground on the end, or it may have an insert or tip attached to it to provide a harder cutting surface. The insert may be a piece of high-quality tool steel or a special material such as sintered carbide and be brazed, welded, or fastened to the tool shank by mechanical means.

Tools may be sharpened or shaped by grinding, and inserts may be replaced or rotated when chipped or worn. Frequently a groove or irregularity is ground in the face of the tool behind the cutting edge. This groove, called a chipbreaker, causes removed stock to break into chips or small curls thus becoming easier to manage or to dispose of.

Multipoint tool. A second type of cutting device commonly used on such power tools as milling machines, horizontal boring machines, and planer millers is the revolving multitoothed cutter (Fig. 3). This multiple cutting edge tool gives the advantage of much faster cutting than can be obtained using a single-point tool. Like the single-point tool, each cutter tooth has a rake angle and clearance angle. The rake angle aids in chip removal and in most cases provides a sharper cutting edge. The clearance angle allows the tooth area behind the cutting edge to clear the work, avoiding drag or friction.

tion. Probably the most common ... regular or plain milling cutter. This cylindrical type of cutter is made in varying diameters and face widths. Cylindrical cutters are made in face widths that range from the thin metal slitting saws to slab mills several inches in width. Side milling cylindrical cutters have teeth on one or both sides.

Other cutters may be constructed in the form of a cone to mill angles or to do beveling.

The face-milling cutter, another frequently used tool, has teeth arranged around the edge of a plate or disk. Usually several inches in diameter, these cutters are able to face or mill a large area rapidly by traversing along it.

The end mill is a cylindrical cutter with teeth on its circumferential face and one end. An integral shank on the other end is used for holding and driving. End mills vary in size from a fraction of an inch to several inches in diameter. Regular milling cutters are able to traverse mill at a depth thus producing a slot or groove. Two-lip end mills are able to cut directly into the work in much the same manner as a drill.

Speed and feed. Two important considerations are present in the relative movement of the cutting tool and workpiece: cutting speed and feed. The cutting speed is the relative surface speed between the work and the tool. For example, in a turning operation, cutting speed is usually measured on the uncut surface of the work ahead of the tool. Cutting speed is commonly given in feet per minute. The feed is the relative amount of motion or travel of the cutting tool into the workpiece per revolution, stroke, or unit of time.

To obtain high machining efficiency, desired surface finish, and appreciable tool life, speed and feed must be closely correlated during any machining operation. Many factors affect this relationship including the type of tool used, its durability, and the rigidity of the machine and tool, as well as the power of the machine, and the material, hardness, and configuration of the workpiece. No general rule

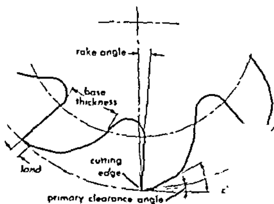


Fig. 3. Typical milling cutter teeth.

or equation for the correct feed and speed can cover all conditions. Some information comes from publications and handbooks, some from the advice of others, and much from personal experience.

Most manufacturers of machines and cutting tools supply information on recommended feeds and speeds for their particular product. In addition, machine shop handbooks provide adequate information on feeds and speeds for normal conditions encountered in machining.

Cutting fluid. Most metal-machining operations require the use of some type of cutting fluid. Machining of metals causes deformation, rubbing, and friction. The resulting temperature rise can warp the work and damage or excessively wear the tool. Also, the metal will tend to expand, causing inaccuracies in the work. Cutting fluids are used mainly for cooling, but they may serve other purposes. In certain instances the surface finish is improved by the use of a cutting fluid. Often the fluid also serves to lubricate the slides on the machine or to protect the machine from corrosion. In many operations the fluid washes away metal chips and particles that could clog or interfere with the tool and the machine.

There are two general types of cutting oils, one based on mixtures with water, the other on oil compounds. In general, if cooling is the primary requirement, an oil-and-water emulsion is used. If lubrication of slides or other parts is required in addition to cooling, a mineral oil compound is usually employed. There are, however, no hard and fast rules and in many cases only experience or actual test will tell the type of cutting oil to use. *See* BORING; BROACHING; DRILLING MACHINE; GEAR CUTTING; LATHE; MILLING MACHINE; NIBBLING; PLANER; REAMER; SAWING; SHAPER; THREADING. *See also* GRINDING; LAPPING; POLISHING. [A.T.]

Mackerel

Any of several species of fishes of the family Scombridae, all of which live on the high seas and are wide-ranging species; a few are cosmopolitan. The common mackerel, *Scomber scombrus*, is the best known. This North Atlantic fish is an excellent food animal, with firm, oily flesh. It is a wanderer, traveling in large schools, and approaching the American shores first near Cape Hatteras in the spring, arriving northward later, and returning to the deep sea in the fall and winter. Mackerel are carnivorous, eating smaller fishes and crustaceans. The



The mackerel, *Scomber scombrus*; length to 2 ft. (*From* E. L. Palmer, *Fieldbook of Natural History*, *McGraw-Hill*, 1949)

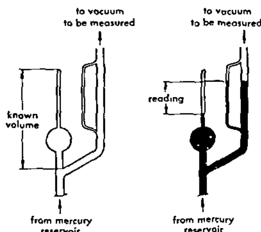
common mackerel is metallic blue above, marked with wavy dark lines, and white below. It is about 2 ft long when mature. This is an important commercial food fish. The annual harvest by American fishermen varies from 4,000,000 to 10,000,000 lb in recent years; formerly the harvest was much greater.

The Pacific mackerel, *Pneumatophorus diego*, a similar species, is even more abundant than the common mackerel. In 1936 the harvest of this species was 146,000,000 lb from California waters. In recent years the catch has been less than one-fifth that amount. *See* PERCIFORMES. [J.D.B.]

McLeod gage

A type of instrument used to measure vacuum. *See* VACUUM MEASUREMENT.

A known volume of the gas whose pressure is to be measured is trapped by raising the level of fluid (almost invariably mercury) by means of a plunger, by lifting the reservoir, by a pressure bulb, or by tipping the apparatus. As the mercury level is further raised, the gas is compressed into the capillary tube. Obeying Boyle's law, the compressed gas now exerts enough pressure to support a column of mercury high enough to read. Readings are very nearly independent of the composition of the gas.



McLeod gage.

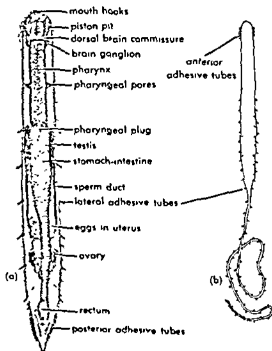
The McLeod gage is simple and inexpensive. It is widely used as an absolute standard of pressure in the range of 0.0001–10 millimeters of mercury; it is often used to calibrate other vacuum instruments which are more convenient to use.

The McLeod gage suffers the disadvantages that readings are discontinuous, that a certain amount of hand manipulation is necessary to make a reading, and that the reading is visual. Mercury vapor may cause trouble by diffusing into the vacuum being measured.

Condensing vapors cannot be measured with a McLeod gage, although an absorber or cold trap may permit measurement of partially condensable mixtures. [B.D.H.; H.C.P.]

Macrodasyoidea

An order of the class Gastrotricha in which the adhesive tubes are distinctive. There are as many as 250 of these structures. They are cylindrical, cuticular projections which are moved by delicate muscle strands. Minute glands, clustered ventrally on the head surface or disposed in lateral or transverse longitudinal rows, produce a sticky secretion. They are part of the syncytial epidermis lying beneath them. Head differentiation is not usual. The sides of the body are parallel, and the shape of the animal is elongate. A few species show anterior



Macrodasyoidea. (a) *Macrodasys*, marine gastrotrich, with lateral adhesive tubes. (b) *Urodasys*, marine gastrotrich, with long tail. (*After A. Remane, 1926, from L. H. Hyman, The Invertebrates, vol. 3, McGraw-Hill, 1951*)

red pigment spots, possibly light-sensitive. Protonephridia for elimination of metabolic waste are lacking. Hermaphroditism occurs in members of this group. *Macrodasys*, *Turbanella*, and *Thaumastoderma* are representative genera, comprising numerous Old World species. A few are now being reported from the United States with unpublished data known to exist. *See* GASTROTRICHA. [C.E.P.]

Macroevolution

The larger course of evolution by which the categories of animal and plant classification above the species have been evolved from each other and have differentiated into the forms within each. Only that part of evolution for which concrete evidence exists will be considered.

The organisms observed, alive or as fossils, can be classified into large groups or phyla. Within each of these the body is organized in variations of a single fundamental type. Such groups are the annelids, mollusks, echinoderms, and other phyla. Macroevolution is the process by which evolution has taken place within these phyla. In some cases evidence can be given that two or more such phyla have earlier evolved from a single ancestral group.

The evidence for the reality of macroevolution is derived from many sources. Paleontology provides evidence of the course of evolution from fossils. Fossils preserved at intervals of geological time may be compared and deductions made of their relationships. Thus, phylogenetic trees may be constructed. Studies of comparative morphology and embryology give evidence of many kinds from the similarities and dissimilarities of living animals. Geographical distribution of forms, together with the known history of the lands, gives data which should be in agreement with beliefs concerning the evolution of the organisms.

Speciation or microevolution has continued throughout evolution and provides its background (*see* SPECIATION). But the question remains whether the course of change in the long progress of macroevolution has been controlled in other ways besides the mechanisms considered in speciation. Another question for discussion is whether macroevolution is merely long-continued speciation. Before the evidence on this question can be considered, some characters of the general nature of macroevolution must be noted.

GENERAL CONSIDERATIONS

Throughout the course of evolution, change has occurred during the life of organisms. These organisms must always be competent to survive in competition with other organisms. The whole biology evolves; evolution is a phenomenon in natural history. Also, change of habit and habitat normally accompanies the evolution of structure. It is probably without meaning to ask which of these precedes the other.

Types of change. In all evolution one may distinguish two types of change, cladogenesis and *phyletic evolution*. *Cladogenesis* is the type of evolution associated with altered habit and habitat, *usually* in parts of a species separated from the rest. Adaptation to the new conditions results in division of the originally homogeneous species into distinguishable parts. Phyletic evolution is the gradual evolution of a population in its environment without division into isolated parts. This consists partly in improved adaptation to the conditions, partly in readaptation when the conditions change, as they frequently do in natural environments. Usually change in environmental conditions is gradual, but sudden and irreversible change sometimes occurs and organisms often spread into new

habitats where conditions differ from those to which they are adapted. The organisms must then adapt quickly to the new conditions if they are to survive. In such circumstances evolution will be rapid under the changed selection due to the new conditions. This process is known as quantum evolution and is considered as a special although extreme case of phyletic evolution. This type of evolution may lead to new groups on any taxonomic level. The distinctions between these types of evolution are illustrated diagrammatically in Fig. 1. The evolutionary changes in the organisms are similar in all the types.

Adaptive radiation. The course of evolution differs between successful groups, such as the vertebrates or the insects, which have enlarged their areas of dominance with time and evolved into many smaller groups, and other groups, such as many forms among the lower invertebrates, which show no progressive evolution and have often survived through long ages with little change.

The knowledge of vertebrate evolution is the most detailed. When the evolution of this successful group is examined it is found that a limited number of dominant types (several groups of fishes, amphibians, reptiles, birds, and mammals) have followed each other, each in a new mode of vertebrate life, and that each was evolved from the previous dominant group. Further, each dominant group divided soon after its appearance into a large number of subsidiary types adapted to more restricted modes of life within the range of the larger group. This diversification into different adaptive zones is known as adaptive radiation. The radiation of the primates, the mammalian order including the monkeys, apes, and man, is illustrated in the diagram in Fig. 2. In adaptive radiation the greater part of the changes of form are due to alterations of the relative sizes of the parts of the body by modifications of their growth rates during development and not to the evolution of new organs. In the mammals the radiating lines appear today as the orders of bats, primates, rodents, and so on. All of these can be traced back to near the beginning of the dominance of the mammals. Each line continues to radiate throughout its successful history. Formation of new organs, which are not modifications of organs previously present, occurs in the radiation but accounts for only a small part of the changes. The development of horns in the mammals is an example.

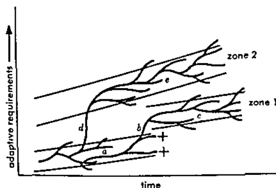


Fig 1. Diagram of the main types of evolution. (*a*) Phyletic evolution in an adaptive zone or niche (zone 1). (*b*) Sudden alteration of conditions in this zone with death (+) of most of the lines inhabiting it; readaptation of one line to the new conditions—quantum evolution. (*c*) Phyletic evolution in the new conditions. (*d*) Migration to a new niche (zone 2 cladogenesis). (*e*) Phyletic evolution in this zone. Rapid (tachytelic) evolution at (*b*) and (*d*); horotelic evolution at (*a*), (*c*), (*e*). (*Modified from G. G. Simpson, Tempo and Mode in Evolution, Columbia Univ. Press, 1944*)

The evolution of a major dominant group from a preceding group, for example the mammals or birds from the reptiles, or the amphibians from the fishes, demands much more fundamental changes than those of adaptive radiation. The whole form and function of the body is reorganized. It takes a considerable time, for example about 40,000,000 years in the mammals and birds, and proceeds by successive alterations of one feature after another toward the new type. It occurs in one of the radiating lines of the preceding group, while its other lines, though they continue to radiate, clearly remain members of the preceding group.

SPECIAL PROBLEMS

Some of the more special problems raised by the study of macroevolution are considered next.

Adaptation. Adaptation is the process by which, as the result of selection, the biology of an organism becomes capable of satisfying the needs of its life. It includes besides the general adaptation that all organisms must possess to be viable, many more special forms such as mimicry and protective resemblance. The conditions to which adaptation is necessary are of many kinds. The physical conditions of its surroundings, the biotic conditions of the organism's contacts with other species, the intraspecific conditions of its contacts within its own species, and the conditions within the organism's body needed to maintain its life and health, all require adaptation (*see* PROTECTIVE COLORATION). If distantly related animals live similar lives under like conditions, their adaptations may be similar, and they are said to be convergent. Thus the body forms of whales and porpoises are convergent with those of many fishes and ichthyosaurs (Fig. 3). [illegible]

webbed feet for swimming, and the parallel adaptations of marsupials and placental mammals to various habits of life are well known. In parallel evolution related organisms evolve in the same di rection by similar genetic changes; in convergence, organs originally unlike come to resem- [illegible]

likely to be identical.

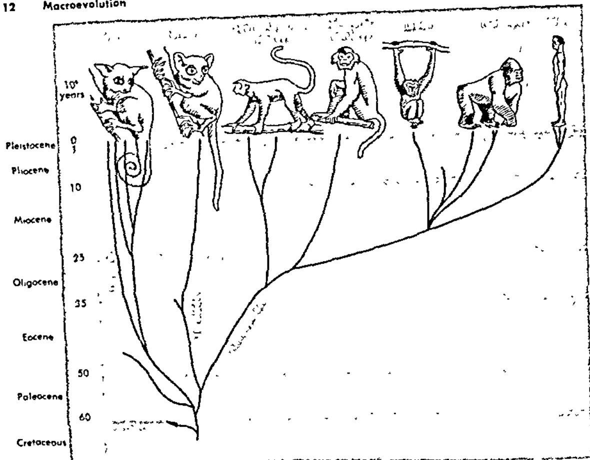


Fig 2 Evolution of the primates. (*Modified from J. Z. Young and P. P. Grassé*)

Specialization and survival. Organisms are often adapted to narrow ranges of conditions in small environments, known as niches, and may be confined to these conditions. Though natural environments are continually changing and change of conditions requires constant readaptation, this specialization may be maintained indefinitely, provided the niche remains available. This close specialization will be of selective advantage, for the organism so specialized is likely to be more efficient in its niche. Its specialization is therefore likely to become closer with time. But specialized forms are always open to the danger that as conditions change their niches may cease to exist and the organisms may be exterminated. Close specialization is therefore a trap into which organisms may be led by the force of selection. Probably for these reasons, new successful groups have generally been evolved from the less specialized members of earlier groups, and the forms which have survived longest have been those adapted to a wide range of conditions.

Preadaptation. In spite of the dangers of specialization, animals specialized to one mode of life have sometimes in evolution passed to another mode of life and become readapted to it. For instance, horses at first lived in forests browsing on leaves and later took to grazing on grassy plains; the fishes ancestral to the amphibians passed from aquatic to amphibious life; many pests have taken to feeding on crops which may be very different from their natural food. Large changes of these kinds probably took place in many steps, but even so they raise the difficulty of explaining how animals specialized to one mode of life have been able to survive in a life in which different adaptations would seem to be required. It is thought that the large changes are made possible when, and perhaps only when, adaptations already made before the change happened to be adaptive after the change, that is, when the animals were preadapted to the change. Since this correlation between conditions before and after the change could only occur by chance, it is unlikely to be frequent, and, in fact, large changes of this type are not frequent in evolution.

Correlation. In order that the organism may remain viable and efficient enough to survive under competition, it is necessary that all parts of the body should be so correlated in structure and activities as to be able to work together as a single, efficient whole. This correlation must always be maintained in evolution. Changes which may be advantageous for the functioning of one part may be disadvantageous to the functioning of the whole body unless other parts are modified. Thus, in mammals large horns may be valuable in fighting but they require strengthening of the bones of the

skull and of the neck muscles to support them; elongation of the horse's legs required simultaneous lengthening of the neck to enable it to graze. Where the change in an organ is small, correlation may be maintained by the required changes in other organs taking place during the individual's life history, in response to the demands made on these organs, for example, as the blacksmith's arm becomes stronger with use. This principle is known as organic selection. Such phenotypic changes will not be inherited, and therefore will not be of evolutionary value, but they may allow the organism to survive until mutations giving the required result occur. If the functional change is large and sudden, organic selection will not be able to maintain correlation. This is probably one of the reasons why large and sudden changes are at least rare in evolution.

The need to maintain correlation must reduce the speed of evolutionary change, but since variation occurs in all parts of the body it need not prevent it. It implies that evolution should be regarded as a process of change in the whole organization of the body, not in its parts separately.

Hypertely. In numerous animals organs are found so fantastically overdeveloped that it is hard to believe that they are not harmful to the activities of the animal. Such organs are called hypertelic. The peacock's tail and the horns of the Irish elk are examples among many. These organs are not useless; that is, they have their uses in sexual activities, fighting, or other activities, and the survival of the animals, now or in the past, shows that they are not so harmful as to render the animals nonviable. The organs are specializations to definite functions in

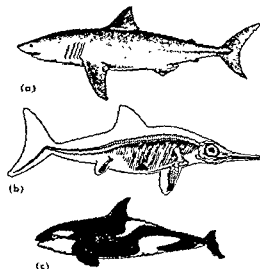


Fig. 3. Convergent evolution in the form of the body in three actively swimming marine vertebrates. (a) A shark (from J. Z. Young, *The Life of Vertebrates*, Oxford Univ. Press; Clarendon Press, 1950). (b) An ichthyosaur (Reptilia). (c) The killer-whale (*Orcinus*, Mammalia) (from J. Z. Young, *The Life of Vertebrates*, Oxford Univ. Press; Clarendon Press, 1950).

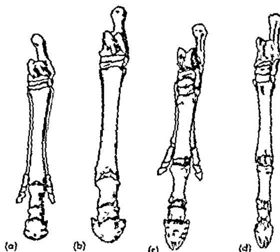


Fig. 4. Parallel evolution in the reduction of the digits in the horses and the Litopterna, a South American group of mammals. (a) Three-toed horse (*Protohippus*). (b) Single-toed horse (*Equus*). (c) Three-toed litopternan (*Diadiaphorus*) (d) Single-toed litopternan (*Thoatherium*). (From G. S. Carter, *Animal Evolution*, Sidgwick and Jackson, 1951)

the life of the animals, and as such their development will be favored by selection until they so far impair viability as to lead to extinction. In some cases they may be genetically bound to other and valuable characters. In deer the relative size of horns increases with body size. If large size is valuable to the elk, the horns should increase in size beyond the size that is optimum for them. A balance would be struck between too large horns and the advantage of large size

Irreversibility. It is often said that evolution is irreversible. This is an overstatement, for often organs previously evolved have been lost; for example, snakes have lost their legs and birds the reptilian teeth. Their evolution has thus been reversed. But it is true that when an organ has been lost it is hardly ever regained in later evolution in the same form. This rule, proposed by L. Dollo, [illegible] an immediate inference from the fact that the genetic backgrounds of all the organs are [illegible] many genes determining the structure and [illegible] tions of each. If this complex of genes [illegible] not likely to be exactly regained by [illegible] A new organ evolved to serve the same purpose will differ in its details. Dollo's rule is [illegible] acters controlled by single genes [illegible] by a few genes. These may be [illegible] tation.

Extinction. The fauna [illegible] change with time either by [illegible] earlier forms or by their [illegible] isms rarely survive indefinitely. [illegible] due to various causes. [illegible] have caused the extinction [illegible] soon after their evolution [illegible] changed conditions [illegible]

eral cause; the extinction of many groups of reptiles in the Late Cretaceous and of many other groups in the Permian was probably caused by changes of climatic conditions. Competition with forms later evolved has very often led to the extinction of earlier types. In general, evolution has progressed toward better organization and greater efficiency. It is to be expected that the earlier forms should most frequently be eliminated.

In view of the progressive character of evolution it is at first sight surprising that organisms are found alive today on all levels of organization from the simplest unicellular forms upward. Many of the simpler organisms are common and widespread and even dominant in restricted habitats. One might expect them to have been exterminated by organisms later evolved. The explanation seems to be that these simple forms are well adapted to the life they lead and either are able to avoid the competition of more complex forms or, like many coelenterates, have evolved organs of defense against their attacks. In some cases they have been able to hold their own by adaptation to special habits of life such as parasitism or other forms of association with the later forms.

Relicts. Besides the simple organisms that are still widespread, a few representatives are often found of groups that were once dominant but have become otherwise extinct. The lampreys and hagfishes alone represent a very early vertebrate group, the Agnatha; the king crab, *Limulus*, is the only representative of a formerly widespread group of arachnids; *Peripatus* is the only living type of the Onychophora, an early arthropod group. These relicts are forms that have evolved slowly, and have escaped extinction by close adaptation to the conditions of their life. Geographical *relicts* are also found surviving in restricted regions where they have found refuge. The shrimp *Mysis relicta* is found in lakes of northern Europe and is probably a marine form isolated when the lakes were closed and later adapted to fresh-water life; *Peripatus* occurs today in several forest regions of the warmer countries, these being probably relict parts of a much wider distribution.

Rates of evolution. The rates at which organisms evolve vary from almost zero in a few forms that have survived as genera from the Paleozoic to Recent times (*Lingula*, Brachiopoda, Cambrian to Recent, 450,000,000 years; *Ostrea*, Lamellibranchiata, Carboniferous to Recent, 250,000,000 years) to the rapid evolution of the char (*Salvelinus*), a fish that has evolved to at least subspecific level in European lakes in the 12,000 years since the end of the Ice Age. Even more rapidly, the Faeroe house mouse has evolved a new subspecies in historic times, in the last 500 years.

Between these extremes lie the great majority of evolutionary rates, but these also vary widely. F. E. Zeuner has estimated that the normal rate in the mammals, a progressive group, is for a *specific difference* to evolve in 1,000,000 years and a generic degree in 4,000,000 or 5,000,000 years. Many groups evolve more slowly. G. G. Simpson in 1953 found that the mean survival time for genera of bivalves, the Lamellibranchiata, is about 50,000,000 years. The rate may vary during the history of a group. Often it is rapid when the group is radiating actively soon after its appearance; later it settles down to a much slower rate.

Both the slow (bradytelic) and rapid (tachytelic) exceptional rates seem to differ fundamentally in their causes from the normal or horotelic rates. It is thought that the differences are not due to different mutation rates. Probably bradytelic rates occur in forms which are almost perfectly adapted to modes of life in abnormally constant environments, so that any change is likely to be harmful. Tachytelic evolution is probably from rapid change in the environment, leading to changes in the direction of selection. In some cases, such as that of the char in European lakes, it is probable that isolation has led to independent evolution and has encouraged differentiation.

One cannot expect to reach so complete an interpretation of macroevolution as of speciation, for the evidence is less complete. But the phenomena observed in it never openly conflict with the modern theory of the causes of evolution. [G.S.C.]

Bibliography: G. S. Carter, *Animal Evolution*, 1951; W. K. Gregory, *Evolution Emerging*, 2 vols., 1951; J. S. Huxley, *Evolution, the Modern Synthesis*, 1943; J. S. Huxley, *Problems of Relative Growth*, 1932; G. L. Jepson, E. Mayr, G. G. Simpson (eds.), *Genetics, Palaeontology and Evolution*, 1949; G. G. Simpson, *The Major Features of Evolution*, 1953; F. E. Zeuner, *Dating the Past*, 2d ed., 1950.

Madelung constant

A numerical constant α_M in terms of which the electrostatic energy U of a *three-dimensional* periodic crystal lattice of positive and *negative point* charges q_+, $-q_-$, N in number, is given by

$$U = -\frac{1}{2}\frac{Nq_+q_-}{d}\alpha_M \qquad (1)$$

where d is the nearest neighbor distance between positive and negative charges and N is large. Knowledge of such electrostatic energies as given by the Madelung constant is of importance in the calculation of the cohesive energies of ionic crystals and in many other problems in the physics of solids. *See* IONIC CRYSTALS.

Designate the lattice sites by indices, i or j, and let the distance between sites i and j be given by r_{ij}. Then U is given by the sum of the Coulomb interaction energies of all pairs of charges:

$$U = \frac{1}{2}\sum_{i,j}{}' \frac{q_iq_j}{r_{ij}} \qquad (2)$$

In the summation, i and j range over all sites in the lattice, and the prime on the summation sign indicates the exclusion of terms for which $i = j$. The factor ½ avoids counting pairs of charges twice.

Madelung constants for some common ionic crystals

Crystal structure	Madelung constant, α_M
Sodium chloride, NaCl	1.7476
Cesium chloride, CsCl	1.7627
Zinc blende, α-ZnS	1.6381
Wurtzite, β-ZnS	1.641
Fluorite, CaF_2	5.0388
Cuprite, CuO_2	4.1155
Rutile, TiO_2	4.816
Anatase, TiO_2	4.800
Corundum, Al_2O_3	25.0312

In Eq. (2), $q_i = q_+$ or q_- depending upon whether the lattice site i is occupied by a positive or a negative ion. From this expression for U it can be seen by comparison of (2) with (1) that

$$\alpha_M = -\frac{1}{2}\sum_{i,j}{}' \left(\frac{q_i}{q_+}\right)\left(\frac{q_j}{q_-}\right)\left(\frac{d}{r_{ij}}\right) \qquad (3)$$

This number is called the Madelung constant after E. Madelung, who first calculated the sum in connection with the cohesive energies of ionic crystals. It is characteristic of the lattice structure but independent of the dimensions of the lattice. Its numerical value does, however, depend on the choice of d. Here d has been chosen to be the nearest neighbor positive-negative charge separation. Occasionally other choices for d, such as the cube root of the molecular volume, or the unit cube edge in cubic crystals, are used instead.

The Madelung constants for a number of common ionic crystal structures are given in the table, in which d has been chosen as the nearest cation-anion distance. See CRYSTAL STRUCTURE. [D.C.D.]

Bibliography: See IONIC CRYSTALS.

Magellanic Clouds

The nearest galaxies to the galactic system. Both the Large and Small Magellanic Clouds (LMC and SMC) are classed as Irregular in the sequence of classification developed by Edwin P. Hubble (see GALAXY, EXTERNAL). Both clouds are visible to the unaided eye from the Southern Hemisphere. Their apparent diameters on the plane of the sky are 7° and 4°, respectively. Because of their nearness, the clouds are of great importance in studies of the stellar content of galaxies, especially in determining the relation between the intrinsic luminosities of various types of stars such as Cepheid variable stars, bright B and A stars, red supergiants, and normal novae. These particular objects are used as distance indicators for more distant galaxies (see VARIABLE STAR).

The period-luminosity relation for Cepheid variable stars has recently been obtained with a magnitude sequence extended to cover Cepheids of periods as small as 2 days. This work refined the slope of the relation between period and magnitude. The correction affects to some extent the determination of all distances to external galaxies and shows that the distance to both of the Magellanic Clouds is the same and is about 60,000 parsec.

The rotation of the clouds was first detected by optical radial velocity measurements of B and A supergiant stars and H II regions, and later by more extensive radio observations using the 21-cm line of neutral hydrogen. The rotational speed of the Large Magellanic Cloud given by the optical data is higher by a factor of 2 than that given by the 21-cm radio data. This difference is believed to show a variation in the rotational motion of the Large Magellanic Cloud between points lying in and points lying above the fundamental plane. The rotational curve for the optical data shows the velocity about the center to be about 40 km/sec for points 4° from the center of the Large Cloud (4° equals 4000 parsecs at a distance of 60,000 parsecs). The systematic motion of the center of mass of each cloud with respect to the center of our galaxy is about +35 km/sec for the LMC and −15 km/sec for the SMC in the radial direction. In addition to these motions, each cloud shows internal random motions in the order of 20 km/sec for the hydrogen gas. The evidence comes from the width of the 21-cm radio line of neutral hydrogen. [A.R.S.]

Magma

The molten rock material from which igneous rocks are formed. For convenience, molten rock material below the earth's surface will be referred to as magma and that above the surface as lava. Magmas and lavas may not be completely liquid; they may carry abundant crystals in suspension. All gradations exist, therefore, between completely liquid magma and solid igneous rocks. Magma may be considered a mutual solution of silicates and more or less dissolved gases or volatiles. When present in abundance these volatiles act as fluxes and reduce the temperature necessary to keep the material molten.

Magmas that are deeply buried may retain abundant volatiles due to high confining pressures. When erupted to or near the surface, the pressure is reduced; and dissolved gases may separate to form bubbles and escape. Except for these fugitive constituents, the composition of most magmas may be closely represented by chemical analyses of carefully selected igneous rocks. See IGNEOUS ROCKS.

Temperatures of magmas. Magmas probably crystallize within wide temperature ranges. They may begin to crystallize around 1100°C and become completely solid before reaching about 600°C. This range is deduced in part from thermal characteristics of minerals and mineral associations formed from magmas. Changes in the rocks (baking effects) in contact with magmas furnish additional information. Freshly erupted Hawaiian lava has a temperature of roughly 1100°C. Lavas generally crystallize within a higher temperature range than do deeper magmas because they lose volatiles which function as fluxes. As will be shown, granitic magmas crystallize within a lower temperature range than basaltic ones. See AUREOLE, CONTACT.

Viscosity. This property is difficult to measure in lavas. It increases with drop in temperature a

is higher in silica-rich lavas (rhyolitic) than in basaltic ones. One might expect silica-rich magma (granitic), therefore, to be more viscous than silica-poor magma (gabbroic), but the compositional difference is offset, in part at least, by the higher volatile content of granitic magma.

Crystallization. As temperature falls, due to radiation to surrounding rocks, solidification sets in. Rapid cooling reduces the activity of the ions to such a degree that they freeze in their highly disorganized arrangement to form glass. This occurs where hot lava cools rapidly at the surface to form obsidian, and where magma, filling narrow fissures, is chilled as it comes into contact with the cool walls. *See* Obsidian.

Slower cooling permits time for grouping of ions to form structural arrangements which are stable at lower temperatures. This is crystallization and it begins first with the generation of tiny nuclei. Some of these are immediately destroyed by ionic bombardment; others manage to enlarge themselves faster than they can be broken asunder and may develop into microscopic crystals (microlites) or into mineral grains many centimeters across.

Crystal growth rate. The growth of crystals depends upon many factors. Each mineral has its own specific power of crystal growth, a property which may vary with temperature, pressure, and other conditions. High viscosity in the magma retards growth, but increase in concentration promotes it. From chemical studies, it seems probable that trace amounts of certain impurities greatly affect crystal growth. In general, crystal growth is slow at first; then it increases to a maximum and falls off rapidly.

Crystal birth rate. Initial crystallization in a magma is believed to depend upon the rate of cooling. In a motionless magma free of inoculating crystals, considerable undercooling *may be required* to initiate crystallization. Most magmas, however, are more or less agitated and have incorporated crystals from their walls or accidentally included rock fragments. Consequently, crystallization may set in at the *freezing point where* the potential birth rate is very low. With greater supercooling the potential increases to a maximum and then falls off to zero. Thus, a deeply buried magma, which loses its heat slowly and is unlikely to be more than slightly supercooled, *gives birth to* relatively few crystals and solidifies with a coarse grain. A lava flow, however, may be quickly cooled (well below its crystallizing temperature) and give rise to many crystals and a very fine grain. Extremely rapid cooling (chilling), as at the surface of a lava flow, is likely to produce more or less glass. As a rule lava flows, dikes, and sills are coarser grained at their interiors because of the difference in cooling rate.

Order of mineral formation. Minerals crystallize from rock melts in a systematic manner, but the order of formation seems to follow no simple fixed rule. Microscopic study of rock textures helps to determine the mineral sequence for specific rocks, and much information can be gathered from relations of rocks in the field. For example, large well-formed crystals are likely to have formed earlier than the finer-grained matrix in which they lie. Interstitial glass and tiny grains in basalt are younger than the crystals of plagioclase forming the interstices. Minerals formed at the margin of a dike are generally younger than those formed near the center. Early crystals are apt to show a better crystal outline than later ones.

Phase equilibria studies of silicate systems provide abundant information concerning sequence of crystallization. Two examples serve to illustrate.

In a simple eutectic system the component in excess of the eutectic composition starts crystallizing first (Fig. 1). A liquid at L1 (composition: A, 48%; B, 52%, and temperature T1) may cool to L2 (boundary of two fields at temperature T2) at which time crystals A2 (composition: pure A) form and enrich the liquid in component B. Further cooling to T3 produces additional crystals (A3) and changes the liquid to L3 (richer in component B). Component A continues to separate to T4 when the liquid (L4) has the eutectic composition for the system (the composition with the lowest melting temperature). At T4 crystals B4 (composition: pure B) start to form along with crystals A4 in the eutectic ratio (A, 30%; B, 70%). The liquid is eventually consumed without change in composition or temperature. The resulting mixture of crystals A4 and B4 may *now cool to some temperature* such as T5 without further phase change. Eutectic systems occur in simple mixtures only. Magmas are multi-component mixtures and as such do not form eutectic systems.

In a solid solution series the higher melting component is enriched in the early crystal phase (Fig. 2). A liquid at L1 (composition: A, 25%; B, 75%, and temperature T1) may cool to L2 (boundary of two fields at T2), at which time crystals C2 (composition nearly pure B) form and enrich the liquid in component A. Further cooling causes further crystallization and a continuous reaction between liquid and early crystals, making both richer in component A. At T3 the liquid has changed to L3, the crystals to C3. This change continues with cooling to T4 where the crystals C4 have the same composition as the original liquid

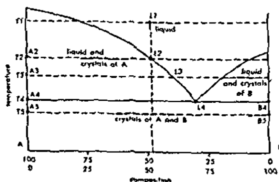


Fig 1. A binary eutectic system.

(L1) and the last drop of liquid (L4) disappears. Crystals C4 (composition: A, 25%; B, 75%) represent a single phase and remain stable as the temperature drops to some point such as T5.

The reaction principle. The principal rock-forming minerals may be placed in four more or less independent divisions: plagioclase feldspars, mafic (iron-magnesium) minerals, quartz, and alkali feldspar.

Plagioclase constitutes a solid-solution series (similar to Fig. 2) in which the calcic member has the higher crystallization point and early crystals are gradually made over into more sodic varieties. Each of the principal mafic minerals (olivine, pyroxene, amphibole, and biotite) belongs to a solid-solution series in which the magnesium member has the higher crystallization point. Furthermore, a significant relationship exists between the four types of mafic minerals. Together they constitute *a discontinuous reaction series in which magma tends to react with olivine to form pyroxene, with pyroxene to form amphibole, and with amphibole to form biotite.* This sequence is controlled largely by temperature, pressure, and water content of the magma. Perhaps somewhat less strikingly, the alkali feldspars show an enrichment in soda over potash as crystallization proceeds. The composition of quartz, however, remains fixed.

Taken as a whole the four divisions are somewhat analogous to the members of a eutectic system since the presence of a member of one division reduces the melting temperature of a member of another division. The relations, however, are more correctly referred to as cotectic. The crystallization of a member of one division may be concurrent with or overlap that of a member of another division throughout a wide range in temperature.

The course of crystallization of the common magmas is very simply expressed in Fig. 3. The early crystals, formed at high temperature (near top of diagram), later become unstable in the cooling magma and tend to react with it to produce lower temperature minerals (farther down in diagram). This process continues until the rock is completely solidified. Quartz is shown at the bottom of the diagram because it generally appears late. Once started, a mineral continues to crystallize until the

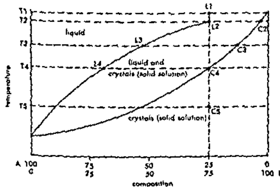


Fig. 2. A solid-solution system.

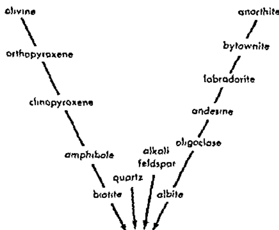


Fig 3. The reaction series.

magma is consumed or until its place is taken by some other mineral or minerals lower in the series.

The diagram shows how higher temperature rocks, such as gabbro, contain mostly olivine, pyroxene, and calcic plagioclase and lower temperature rocks, such as granite, carry biotite, quartz, and alkali feldspar.

There are numerous exceptions to this general picture. Under some conditions a mineral may start to crystallize, later become unstable and be replaced by another, and finally appear again at a still lower temperature. The exact sequence may be affected by the changing composition of the magma and the relative amount of water contained in it. If time does not permit reaction to be completed, unstable relicts (commonly with reaction rims) may remain. Olivine rimmed by enstatite and augite rimmed by hornblende and biotite are common. Incomplete reaction is frequently expressed by strongly zoned crystals of plagioclase.

As magmatic crystallization ceases, a hot residual fluid, rich in water and other volatiles and carrying much silica, alumina, and alkalies, is believed to move out of the igneous rocks to form certain pegmatites and mineral deposits. *See* ORE AND MINERAL DEPOSITS; PEGMATITE.

ORIGIN OF MAGMA

The origin of magma constitutes a major problem for which various theories have been suggested. Some magmas may be primary in the sense that they represent initially molten material within the earth. Others may have formed from time to time by local melting of deeply buried rock. Downthrusting or folding in zones of mountain building might bring cold rock within sufficiently hot regions of the earth for melting to take place.

Earth structure and primary magma. Geological and geophysical evidence indicates that the earth, in its outer part at least, is composed of layers or shells of different material. From the outermost, granitic layer one may pass downward through successively heavier layers to basaltic and peridotitic material.

The vast preponderance of basaltic rocks, erupted as volcanic material, lends support to the idea that basaltic magma is primarily derived from a deep earth shell of great areal extent. Such magma might be derived by complete melting in the basaltic layer or by selective melting in the underlying peridotitic shell. The abundance of granitic rocks in zones of fold-mountains suggests that granite magma may also be formed in great quantities by selective melting of granitic or intermediate shells. From two such magmas (basaltic and granitic) may come most igneous rocks. Other magmas (such as anorthositic and peridotitic), however, have been postulated to account for some of the less common rock types.

Development of secondary magmas. Magmas derived from earlier melts subsequent to eruption from deep earth reservoirs may be of various compositions and origins. They may form by assimilation of foreign material, by differentiation, or by mingling of more primary magmas.

Magmatic assimilation. The assimilation of foreign material by magmas probably plays a major role in the generation of secondary magmas. The composition of a hot melt may be radically changed by incorporation and digestion of much foreign material.

It follows from the reaction principle that foreign fragments (xenoliths) may not be in equilibrium with the enclosing melt. Certain minerals in the xenoliths may be equivalent to those higher in the reaction series than the minerals actually precipitating from the magma. These will react and be made over into minerals in equilibrium with the magma. Xenolithic constituents which are lower in the series than the minerals actually precipitating from the melt will be dissolved. Assimilation of this type may have been responsible for the formation of certain diorites. Assimilation of limestone is believed to account for the formation of certain nepheline-bearing rocks. Assimilation of vast quantities of sialic material (sediments, old granites, etc.) by primary basaltic magma may help to explain the abundance of granites in some areas.

Magmatic differentiation. Differentiation of a more primary silicate melt could lead to a magma or magmas (and ultimately rocks) of different composition. It is apparent from the reaction principle that as crystallization proceeds the residual magma becomes enriched in silica and alkalies and relatively rich in iron as compared with magnesium. A primary basaltic (or gabbroic) magma starts crystallizing olivine and calcic plagioclase, and the residual liquid becomes more granitic. If these early crystals could be extracted from the melt at this stage by gravity or by squeezing out of the liquid, a magma of dioritic composition would be formed. If continued, the process might lead to the formation of granitic magma. This means of differentiation is called fractional crystallization and appears clearly to have operated in the formation of certain sheetlike masses of gabbroic rock. *See* DIABASE; GABBRO.

Gaseous transfer of material in a magma may assist differentiation. Gas bubbles may effectively transport chemical constituents from lower to higher levels in a magma chamber, or they may remove them completely by escaping through fractures in the overlying rocks.

Liquid immiscibility in magma may play a role in differentiation. Supposedly upon reaching a certain temperature, a homogeneous melt would separate into two complementary liquid phases with droplets of one within the other. Density differences between the two phases might permit rather complete separation on a grand scale and the formation of complementary magmas (such as basaltic and rhyolitic). Though many geological relations suggest this possibility, experimental evidence more or less rules it out. This evidence, however, does not eliminate the possibility of immiscible relations at very high temperatures. Liquid phases separated at high temperatures could mix again at lower temperatures to form magmas of different, intermediate compositions.

Mingling of magmas. The mixing of magmas, such as granitic and basaltic, might produce secondary magmas of intermediate composition; but the mechanism is not generally considered to have operated on a large scale. Mingling of secondary magmas in closely spaced subvolcanic chambers is believed to be important locally. *See also* PETROGRAPHIC PROVINCE. [C.A.C.N.]

Bibliography: T. F. W. Barth, *Theoretical Petrology*, 1952; N. L. Bowen, *The Evolution of the Igneous Rocks*, reprint, 1956; F. J. Turner and J. Verhoogen, *Igneous and Metamorphic Petrology*, 1951; E. E. Wahlstrom, *Introduction to Theoretical Igneous Petrology*, 1950.

Magnesite

The mineral form of magnesium carbonate, $MgCO_3$. Iron, manganese, and cobalt may replace some magnesium in magnesite. Magnesite has hexagonal (rhombohedral) symmetry and the same structure as calcite. It is usually massive and white but iron impurities may give it a brownish tint. The specific gravity is 3 and the hardness is 4 on Mohs scale. Magnesite is stable up to 740°C at 10,000 psi and up to 850°C at 30,000 psi of carbon dioxide. The equilibrium replacement of magnesium by calcium increases with temperature up to 2% at 900°C.

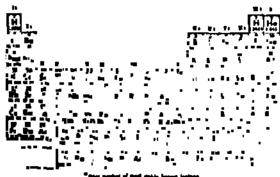
Magnesite may be associated with peridotites, soapstones, and dolomites or it may occur as sedimentary deposits. Deposits are in Austria, Manchuria, California, Nevada, and Washington. Magnesite is a source of magnesia. *See* CARBONATE MINERALS; MAGNESIUM. [R.I.H.]

Magnesium

Chemical element number 12, magnesium, Mg, is a silvery white metal which has become well known as the world's lightest structural metal. The specific gravity of magnesium (1.74) is two-thirds that of the other well-known light metal, aluminum.

Although magnesium is probably best known because of its lightness in structural applications, the

metal has many desirable chemical properties which account for its extensive usage in many nonstructural applications.



Sir Humphry Davy is generally credited with the discovery of magnesium, in 1808, when he established the fact that the compound magnesia alba (magnesium oxide) was the oxide of a new metal. Davy passed potassium vapors over hot magnesium oxide and extracted the reduced magnesium with mercury. He also electrolyzed magnesium sulfate, using mercury as a cathode. In both cases, he obtained magnesium in the form of an amalgam, but it is not known definitely whether he actually obtained metallic magnesium. The first actual isolation of magnesium is attributed to the French scientist A. A-B. Bussy, who in 1830 fused anhydrous magnesium chloride with metallic potassium and obtained magnesium that was substantially pure. Michael Faraday was the first to produce magnesium electrolytically in 1833 by the electrolysis of molten magnesium chloride, using a voltaic cell. Later, in 1852, R. Bunsen designed an electrolytic cell in which hollow carbon cathodes were used to collect the molten magnesium to prevent burning when it came into contact with air.

Natural occurrence. Magnesium is very abundant in nature, occurring in substantial amounts in many rock-forming minerals such as dolomite, magnesite, olivine, and serpentine. In addition, magnesium is also found in sea water, subterranean brines, and salt beds. It has been estimated that magnesium constitutes 2.5% of the earth's crust, making it the sixth most abundant chemical element. Magnesium is the third most abundant structural metal in the earth's crust, being exceeded only by aluminum and iron.

More than 60 minerals occurring in nature contain the element magnesium, but only a few are important commercially for the production of magnesium compounds. In the United States, the important sources are brucite, dolomite, magnesite, natural brines, sea water, and sea-water bitterns. Olivine and serpentine offer a huge potential, but these minerals are not being used extensively for the production of magnesium compounds. Although California, Nevada, and Washington are major sources for brucite, dolomite, and magnesite, huge deposits of magnesium ores are widespread throughout the United States. Large reserves of magnesite are found in Austria, Brazil, Canada, Czechoslovakia, Greece, India, Manchuria, Russia, Venezuela, and Yugoslavia.

Production of the metal. Two production methods of major importance have been used in the United States, the electrolytic and the silicothermic processes. Magnesium is produced from sea water by the electrolytic process. The silicothermic, or ferrosilicon, process uses dolomite as the raw material.

Electrolytic process. The electrolysis of magnesium chloride to yield chlorine and metallic magnesium is the basis of this process. Although magnesite, dolomite, and natural brines have been used as raw materials, the principal source is sea water, which contains about 0.13% magnesium. Because of this content of magnesium and the fact that an economical method for its extraction has been developed, the supply of magnesium is considered limitless. As an example, it has been calculated that if magnesium were extracted from sea water at the rate of 100,000,000 tons/year for 1,000,000 years, the magnesium content of sea water would drop only to 0.12%. The method of extracting magnesium from sea water is shown in Fig. 1. The sea water is pumped into large settling tanks where it is mixed with lime obtained by roasting oyster shells dredged from the ocean bottom. The lime converts the magesium into insoluble magnesium hydroxide (milk of magnesia) which is filtered out. This hydroxide is then treated with hydrochloric acid, obtained from chlorine by reaction with natural gas, to produce magnesium chloride solution. The water is evaporated, and the dry magnesium chloride is fed to the electrolytic cells which break it up into metallic magnesium and chlorine. The chlorine gas is recycled to make hydrochloric acid, and the magnesium metal is poured into ingots. See Electrometallurgy.

Silicothermic process. The ferrosilicon process, although first originated experimentally in Ger-

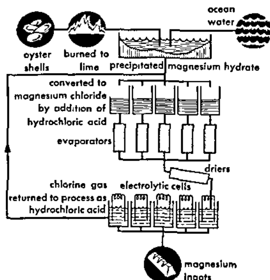


Fig. 1. Diagram showing electrolytic extraction of magnesium from sea water.

many, was developed commercially during World War II by L. M. Pidgeon of Canada. The flow diagram in Fig. 2 shows the actual steps. Ferrosilicon, an alloy of silicon and iron, is mixed with calcined dolomite ore and pressed into small briquets. These are charged into a steel retort, put under vacuum and heated to about 2200°F. The silicon reduces the magnesium oxide (formed by calcining of the dolomite) to form a vapor of metallic magnesium which condenses in the cool end of the retort extending from the furnace. In this process, the magnesium is removed from the retort in the form of crystals which are subsequently melted and cast into ingots.

Properties. The properties of magnesium in metallic form are best divided into two categories, physical and chemical. The former are listed in Table 1. Many of these properties provide magnesium with distinctive qualities that determine some of its uses.

Magnesium is very active chemically, as indicated by its high position in the electromotive force series of metals. It will actually displace hydrogen from boiling water, and a large number of metals can be prepared by thermal reduction of their salts and oxides with magnesium. The metal will combine with most nonmetals and with practically all acids, noted exceptions being pure chromic and hydrofluoric acids. Magnesium reacts only slightly or not at all with most alkalies and many organic chemicals, including hydrocarbons, aldehydes, alcohols, phenols, amines, esters, and most oils. As a catalyst, magnesium is useful for promoting organic condensation, reduction, addition, and dehalogenation reactions. It has long been used for the synthesis of complex and special organic compounds by the well-known Grignard reaction (see GRIGNARD REACTION). Principal alloying ingredients include aluminum, manganese, zirconium, zinc, rare-earth metals, and thorium. Certain combinations of these alloying constituents produce magnesium alloys suitable for sand, permanent-mold, and die castings; extrusions; forgings; and sheet and plate with excellent mechanical properties at room and elevated temperatures. See MAGNESIUM ALLOYS.

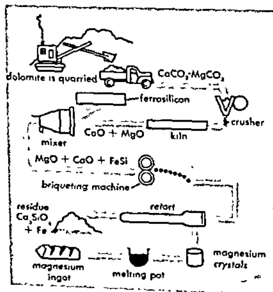


Fig. 2. Diagram showing extraction of magnesium metal from dolomite ore by the silicothermic process.

Table 1. Physical properties of magnesium (99.9% pure)

Atomic weight	24.32
Atomic volume, cm^3/g atom	14 0
Crystal structure	Close-packed hexagonal
Electron arrangement in free atoms	(2) (8) 2
Mass numbers of the isotopes	24, 25, 26
Relative abundances of the isotopes, %	77, 11 5, 11.5
Density at 20°C, g/cm^3	1.74
Melting point, °C	650
Boiling point, °C	1110 ± 10

Workability. The excellent machinability of magnesium is one of its outstanding characteristics. The metal can be machined at higher speeds and with larger feeds and depths of cut than is possible with most other commonly used metals.

Magnesium can be cast and fabricated by practically every metal-forming method known. The metal can be cast in sand or permanent molds and also by the die-casting process which is used when the quantities desired are sufficiently great. Magnesium can also be cast by some of the less-common methods, including plaster, centrifugal, and shell molding, and investment processes.

Magnesium is rolled into sheet and plate, and can be extruded into rods, bars, tubing, and an almost endless variety of structural and special shapes.

Stamping, deep and shallow drawing, blanking, coining, spinning, and impact extrusion are representative of the types of forming operation which can be used to fabricate magnesium.

The forging of magnesium is accomplished by methods much the same as those used for forging other metals. Both press and hammer equipment are used, but the former is most commonly employed because the physical structure of magnesium makes the metal better adapted to the squeezing action of the forging press.

Magnesium parts can be joined by any of the common methods such as arc and electric-resistance welding, adhesive bonding, and riveting. Brazing and gas welding, although not as frequently used as the other methods, are also suitable for joining magnesium. See METAL FORMING.

Uses. Modern developments in the magnesium industry have greatly extended the fields of usefulness for this lightweight metal. There are new alloys with interesting and useful properties. Also, there are new and improved fabricating techniques.

Uses for magnesium alloys can be divided into two types, structural and nonstructural. The structural uses of magnesium are many and varied. During its early years, magnesium was probably best

known for its applications in the aircraft industry, where considerable amounts were utilized in fuselages, engine parts, and accessories. Landing wheels also accounted for large tonnages of the metal. In more recent years, the aspect of unlimited supply and more favorable price position has encouraged many other industries to investigate and use magnesium. As a result, the metal is finding expanded usage throughout industry. The materials-handling field has found magnesium's light weight and durability very desirable for such items as hand trucks, dockboards, grain shovels, gravity conveyors, and foundry equipment. The portable-tool industry is using magnesium castings for drills, chain saws, impact hammers, and all manually handled equipment of this type. Industrial machinery of many types makes use of magnesium's lightness and durability. Among these, textile and printing machines are outstanding examples. Ground-transportation vehicles, including trucks, trailers, and passenger automobiles, offer one of the largest potential uses of magnesium. Household goods, office equipment, instruments of many types, and sporting goods are other examples of well-known applications.

Nonstructural uses of magnesium include its application as an alloying element in aluminum, zinc, and certain other nonferrous alloys. Magnesium is also used as an oxygen scavenger in the manufac-

with 30% and more of aluminum. The cathodic protection of other metals from corrosion is a nonstructural use for magnesium that has become well known. In a similar manner, magnesium functions well in dry-cell battery construction. Gray-iron foundries use magnesium-containing alloys as addition agents to the ladle, just prior to pouring a casting, with the result that the graphite particles become nodular, and the properties of the cast iron are greatly improved. Magnesium, a powerful reducing agent, is used for the production of titanium, zirconium, beryllium, uranium, and hafnium. Because of its rapid yet controlled etching characteristics, as well as its lightness, magnesium finds expanding usage in the photoengraving field.

Principal compounds. Magnesium compounds are used extensively in industry and agriculture. The compounds of major importance are the carbonate, chloride, hydroxide, oxide, and sulfate, which are produced and used in considerable tonnage. Other magnesium compounds, including the bromide, nitrate, phosphate, acetate, silicate, and trisilicate, also find considerable usage in industry, although not to the same degree as the compounds of major importance.

Magnesium compounds are very important to industry. Well-known uses include production of magnesium metal, refractories, insulation, fertilizers, textile processing, leather tanning, paper making, ceramics, explosives, and medicinals. Table 2 lists the major magnesium compounds and indicates some of their more significant applications.

Table 2. Principal magnesium compounds and uses

Compound	Uses
Magnesium carbonate	Refractories, production of other magnesium compounds, water treatment, fertilizers
Magnesium chloride	Cell feed for production of metallic magnesium, oxychloride cements, refrigerating brines, catalyst in organic chemistry, production of other magnesium compounds, flocculating agent, treatment of foliage to prevent fire and resist fire, magnesium melting and welding fluxes
Magnesium hydroxide	Chemical intermediate, alkali, medicinal
Magnesium oxide	Insulation, refractories, oxychloride and oxysulfate cements, fertilizers, rayon-textile processing, water treatment, papermaking, household cleaners, alkali, pharmaceuticals, rubber filler and catalyst
Magnesium sulfate	Leather tanning, paper sizing, oxychloride and oxysulfate cements, rayon delustrant, textile dyeing and printing, medicinal, fertilizer ingredient, livestock-food additive, ceramics, explosives, match manufacture

Several other magnesium compounds, both organic and inorganic, are used in a number of industries for a variety of purposes—as chemical reagents, medicinals, catalysts, and mild abrasives. [W.H.CR.]

Bibliography: A. V. Beck (ed.), *The Technology of Magnesium and Its Alloys*, 3d ed., 1943; W. H. Gross, *The Story of Magnesium*, 1949.

Magnesium alloys

The most important alloying ingredients used in

to 1.83. This low specific gravity has led to a great many structural applications in the aircraft, transportation, materials-handling, and portable-tool and -equipment industries. This article deals exclusively with the structural alloys and uses of magnesium

For a discussion of the history, production, properties, and nonstructural uses of the metal magnesium, see MAGNESIUM.

The first use of magnesium alloys for structural applications was made in Germany, where, as early as 1910, fabricated parts and castings appeared under the trade name of Elektron. Their use increased rapidly because of the local supply of magnesium and the scarcity of other metals during and immediately following World War I. Prior to 1920, the alloys contained only zinc as an alloying constituent. These alloys were displaced during the 1920s and 1930s by development, in both Germany and the United States, of alloys containing aluminum as the primary alloying constituent, with or without the addition of zinc as a third ingredient. All these alloys also contain a few tenths of 1% manganese for added corrosion resistance. Anoth development of this period was the binary

sium-manganese alloy. The use of magnesium in the United States remained at a low level through the 1930s. The great expansion in the use of magnesium came during World War II.

Commercial alloys. The extensive use of magnesium during World War II was based almost entirely on magnesium-aluminum-zinc alloys and, to a much more limited extent, on the binary magnesium-manganese alloy. The Mg-Al-Zn alloys are still the most important. The designations and chemical compositions of the most important of these alloys together with the more recently developed alloys are given in Table 1. Magnesium-aluminum-zinc alloys are produced in the form of sand, permanent-mold, and die castings; extrusions; rolled sheet and plate; and forgings. They are also cast by some of the less-common methods, including plaster mold, centrifugal molding, shell molding, and investment molding processes. Their properties may be modified by appropriate heat treatments. Sand and permanent-mold castings are used primarily in the solution heat-treated condition and in the solution heat-treated plus artificially aged condition. The alloys are used in the solution heat-treated condition, where maximum ductility and toughness with adequate strength are required, and in the artificially aged condition, where maximum strength is needed and somewhat lower ductility can be tolerated. Die castings are used in the as-cast condition. Extrusions and forgings are used either as-fabricated or after an artificial aging treatment. Sheet and plate are produced primarily in two tempers, annealed and strain-hardened plus partially annealed.

An early departure from the Mg-Al-Zn system was made in the late 1940s by the development of magnesium-zinc-zirconium alloys. ZK51A alloy exhibits an excellent combination of strength and ductility in sand and permanent mold castings. ZK60A alloy possesses a similar combination of properties in extrusions and forgings. Both are used in the artificially aged condition. Slabs for rolling and ingots for extrusion and forging of all magnesium alloys are produced by the direct-chill continuous casting process. A significant improvement in the compressive yield strength of ZK60A alloy is realized by producing extrusions of this composition by direct extrusion of atomized pellets (particles of 0.016 in. average diameter) instead of cast ingot.

A sheet and plate composition is ZE10A alloy. Unlike AZ31B, this alloy does not have to be stress relieved after welding. Combined with good strength and excellent weldability and weld efficiency, this advantage offers high productivity in large structural assembly operations.

The alloys described thus far have one disadvantage in that they lose their strength rapidly with increasing temperature, especially above 300°F. The need for alloys with improved strength at elevated temperatures was first met by the development of alloys containing rare-earth metals as the principal alloying constituents. EZ33A is a representative of this family of alloys. The rare-earth metals in this alloy are present as misch metal, the mixture of rare earths occurring naturally in monazite sand. EZ33A is used as sand and permanent-mold castings in the artificially aged state for applications where temperatures in the range 300–450°F are anticipated.

Significantly higher strength properties are obtained in magnesium alloys by the addition of separated rare-earth metals instead of misch metal. The most effective rare-earth metal combination is didy-

Table 1. Chemical composition of common magnesium alloys, %[a]

Alloy	Aluminum	Manganese (minimum)	Rare earth	Thorium	Zinc	Zirconium
AZ31B	3.0	0.2			1.0	
AZ61A	6.5	0.15			1.0	
AZ63A	6.0	0.15			3.0	
AZ81A	7.5	0.13			0.7	
AZ80A	8.5	0.15			0.5	
AZ91B	9.0	0.13			0.7	
AZ91C	8.7	0.13			0.7	
AZ92A	9.0	0.10			2.0	
EK31A			3.0[b]			0.6
EZ33A			3.0[c]		2.7	0.6
HK31A				3.0		0.7
HM21A		0.45		2.0		
HM31A		1.2		3.0		
HZ32A				3.0	2.1	0.7
K1A						0.7
M1A		1.2				
QE22A[d]			2.0[b]			0.7
ZE10A			0.17[c]		1.2	
ZH62A				1.8	5.7	0.7
ZK51A					4.6	0.7
ZK60A					5.7	0.55

[a] Balance is magnesium in all cases. [b] Rare-earth metals present as didymium. [c] Rare-earth metals present as misch metal. [d] Alloy also contains 2.5% silver.

mium, a mixture consisting of about 85% neodymium and 15% praseodymium. This rare-earth metal mixture is used in two sand and permanent-mold casting alloys, QE22A and EK31A. Both are used in the solution heat-treated and artificially aged condition. Complete analyses and property data are included in Table I. The creep resistance of these alloys does not differ markedly from that of EZ33A. Both alloys exhibit good temperature resistance and are important for applications involving exposure at temperatures in the range 300–600°F depending upon the duration of the exposure. In addition, QE22A has the highest yield strength at room and moderately elevated temperatures of any known magnesium casting alloy and accordingly is expected to gain quite wide acceptance as a general composition where maximum performance is required. EK31A also is available as a forging alloy.

The temperature range over which magnesium alloys exhibit structurally useful properties has been markedly extended by the use of thorium as an alloying constituent. Magnesium-thorium alloys are used up to 700–900°F or even higher temperatures, depending upon the duration of exposure at the elevated temperature. The various alloys containing thorium are included in Table I. An important consideration in elevated-temperature applications, especially of long-time duration, is resistance to creep. Magnesium alloys containing thorium and rare-earth metals have high creep resistance. Another important characteristic in structural design is the modulus of elasticity. Magnesium alloys containing thorium and rare-earth metals retain a high modulus over a wide temperature range. The combination of light weight and good strength exhibited by magnesium alloys permits the design of structures with high stiffness-to-weight ratio. Table 2 compares the relative strength and stiffness to bending of magnesium alloys with those of other common structural metals.

A little-known advantage of magnesium over other common structural materials is its inherently high damping capacity, that is, its capacity to absorb mechanical vibrations. In general, addition of alloying elements decreases this property. A new alloy, K1A, has been developed which offers the best combination of damping capacity, strength, and castability. K1A is produced in the form of sand, permanent-mold, and die castings. In all cases, it is used in the as-cast condition.

Alloy preparation. Metal for the production of either shape castings or ingots for subsequent fabrication is prepared by melting and alloying in either stationary or tilting iron pots or crucibles. The capacity of the pots may vary from a few hundred up to 5000 lb. Induction furnaces or ceramic-lined reverberatory or muffle furnaces may also be used. The surface of the molten metal is protected from atmospheric oxidation by a cover of flux consisting of a mixture of alkali and alkaline-earth halides. Aluminum, zinc, thorium, and misch metal are usually added in the elemental form. Manganese and zirconium are alloyed by the addition of a proprietary fused salt containing the chloride of the desired metal which is reduced by the molten magnesium to the metal in place. Zirconium and thorium may be added in the form of magnesium hardeners

Table 2. Relative strengths and stiffness in bending of some structural metals*

Comparison	Material	Thickness	Strength (bending)	Stiffness	Weight
For equal thickness	SAE 1025 steel	100	100	100	100
	6061-T6 aluminum extrusions	100	97.2	34.5	34.5
	AZ31B magnesium extrusions	100	47.2	22.4	22.5
	ZK60A-T5 magnesium extrusions	100	88.9	22.4	22.5
	AZ31B-H24 magnesium sheet	100	73.4	22.4	22.5
	6061-T6 aluminum sheet	100	97.2	34.5	34.5
For equal strength	SAE 1025 steel	100	100	100	100
	6061-T6 aluminum extrusions	101	100	35.8	34.8
	AZ31B magnesium extrusions	146	100	69.2	32.9
	ZK60A-T5 magnesium extrusions	106	100	26.7	23.9
	AZ31B-H24 magnesium sheet	117	100	35.6	26.3
	6061-T6 aluminum sheet	101	100	35.8	34.8
For equal stiffness	SAE 1025 steel	100	100	100	100
	6061-T6 aluminum extrusions	143	199	100	49.3
	AZ31B magnesium extrusions	165	129	100	37.2
	ZK60A-T5 magnesium extrusions	165	212	100	37.2
	AZ31B-H24 magnesium sheet	165	200	100	37.2
	6061-T6 aluminum sheet	143	199	100	49.3
For equal weight	SAE 1025 steel	100	100	100	100
	6061-T6 aluminum extrusions	290	817	841	100
	AZ31B magnesium extrusions	444	930	1962	100
	ZK60A-T5 magnesium extrusions	444	1753	1962	100
	AZ31B-H24 magnesium sheet	444	1451	1962	100
	6061-T6 aluminum sheet	290	817	841	100

* Rectangular beams of constant width with following minimum yield strengths: SAE 1025—36,000 psi; 6061 aluminum—35,000 psi; magnesium alloys—average of minimum tensile yield and compressive yield strengths.

containing 20–35% of the alloying element. Beryllium in amounts of 0.0003–0.001% is added particularly to magnesium die-casting alloys in order to minimize oxidation in the molten state during melting and casting operations.

Magnesium-aluminum-zinc alloys are grain refined either by superheating or by the addition of a carbonaceous material. Alloys containing zirconium are inherently fine grained. Molten magnesium alloys are transferred from one pot to another or from pot to the mold by air-driven mechanical pumps through steel pipes.

Fabricability. In addition to being adaptable to all the primary working operations already mentioned, magnesium alloys can be fabricated by all the common metal-working processes such as stamping, deep and shallow drawing, blanking, coining, spinning, impact extrusion, and forging. For forging operations, both press and hammer equipment are used, but the former is more commonly used because the physical structure of magnesium alloys makes the metals better adapted to the squeezing action of the forging press.

Magnesium alloys exhibit excellent machinability. They can be machined at higher speeds and with larger feeds and depths of cut than is possible with most other commonly used metals.

Magnesium alloy parts can be joined by any of the common methods. Arc and resistance welding, adhesive bonding, and riveting are in daily production use. Brazing and gas welding, although not as frequently used as the other methods, are also suitable ways of joining magnesium alloys.

A wide variety of protective and decorative surface-finishing systems can be applied to magnesium alloys. Magnesium alloys can be treated chemically and electrochemically to produce a protective and paint-adherent surface. In addition, they can be painted, electroplated, anodized, and clad with plastic sheathing.

Uses. The modern developments in alloys, protective surface treatments, fabricating techniques, and primary production facilities have greatly extended the fields of usefulness of the lightweight magnesium alloys. In the early years, the use of magnesium alloys was largely confined to the aircraft industry, where maximum advantage can be taken of their high strength-to-weight relationship. The Mg-Al-Zn alloys have been used in large quantities for engine castings, fuselages, and accessories. Landing wheels also account for large tonnages of these materials. AZ63A-T4 and AZ81A-T4 are the [illegible]

In jet er[illegible]

tempera[illegible]

EZ33A, [illegible]

Use of magnesium alloys in missiles may easily exceed that in piloted aircraft. These alloys have been used in at least 20 different missiles, including ICBMs, and they are being designed into several space projects. [illegible]

of the newly developed alloys HK31A, HM21A, and HM31A. An example of such a missile is the Titan ICBM, which contains 2000 lb of magnesium alloys containing thorium in various product forms.

In terms of sheer weight, size, and production volume, the group of castings for the Nike Hercules tracking and guidance antenna system offer an excellent example. These are produced in AZ91C alloy. A large volume of castings and fabricated parts also goes into the electronic systems of these missiles.

In more recent years, the aspect of unlimited supply and a more favorable price position has encouraged the use of magnesium alloys in many industrial and commercial applications. Large quantities of AZ31B plate in thicknesses varying from 0.250 to 6.000 in. are used for tooling jigs and fixtures. AZ31C alloy sheet in large quantities is now supplied in coils from a modern rolling mill for the manufacture of luggage. Large quantities of magnesium alloy tread plate are finding many uses in dock boards, loading platforms, and mechanical handling equipment. Other applications in the materials-handling field are hand trucks, gravity conveyors, grain shovels, and foundry equipment of many kinds.

The portable tool industry is using magnesium alloy castings for drills, chain saws, impact hammers, and all manually handled equipment of this type. The light weight and durability of magnesium alloys make them useful in many types of industrial machinery. Among these, textile and printing machines are outstanding examples.

Ground transportation vehicles, including trucks, trailers, and passenger cars, offer one of the largest potential uses of magnesium alloys. Extensive use has been made of magnesium alloy sheet and extrusions in the construction of trucks and trailers. Passenger cars have used primarily die castings of AZ91B alloy. A boon to the magnesium die-casting industry has been the development of an automated process for feeding metered quantities of molten alloy into the shot well of a cold-chamber die-casting machine. The economies resulting from this and other recent developments are expected to lead to a greatly expanded use of magnesium in the automotive field. Other important uses of magnesium alloy die castings are electric motor end frames, camera and projector parts, portable tools, electrical connection plugs, tape reels, and missile parts.

Household goods, office equipment, instruments of many types, and sporting goods are other examples of well-known applications in which magnesium alloys are being used. *See* ALLOY; HEAT TREATMENT (METALS AND ALLOYS); METAL FORMING. [T.E.L.]

Magnet

A piece of [illegible]

a net torque when placed in a magnetic field pro-

duced by some other source. For a discussion of ferromagnetic domains, see FERROMAGNETISM.

If a bar of magnetically hard ferromagnetic material, such as Alnico 5, is magnetized to saturation in the magnetic field of a current flowing in a coil of wire, and the current is then turned off, the magnetic flux density B in the bar drops down the hysteresis curve to a point below the retentivity point (see HYSTERESIS, MAGNETIC). The flux density B goes below this point because the field H becomes negative due to the demagnetizing field of the magnet's poles. The important point here is that the bar magnet does retain an appreciable magnetization and, if the material is well chosen, this magnetization will remain more or less permanently, provided the magnet is not heated, dropped, exposed to too strong a magnetic field, or in general, treated too roughly. Such a magnet is called a permanent magnet.

Keeper. This is a piece of soft (easily magnetized) ferromagnetic material (such as soft iron) which (when the magnet is not in use) extends from one pole to the other, and through which the magnetic flux lines between the poles are concentrated. A keeper is used especially with U-shaped or horseshoe magnets. Its presence decreases the demagnetizing effect of the poles and thus helps to keep the magnet strong.

Atomic theory. Modern theory indicates that the magnetic properties of bulk matter can be attributed to the orbital motions and spins of the electrons of the atoms which make up the matter. Orbital motions of electrons and electron spins are the equivalent of tiny electric current loops, and whenever there is a current, a *magnetic field* is always produced. A current loop has a magnetic *moment M* which, by the *Sommerfeld proposal*, is given in mks units by

$$M = IA \qquad (1)$$

where I is the *current in amperes flowing* around the periphery of the area A, the latter being expressed in square meters. The direction of the vector for A (and thus for M) is the direction of the extended thumb of the right hand when the fingers encircle A in the direction of the flow of I. (By the Kennelly proposal in the mks system $M = \mu_0 IA$, where $\mu_0 = 4\pi \times 10^{-7}$ weber/amp-meter is the permeability of free space; see ELECTRICAL UNITS.) The magnetic moment of an atom is the vector sum of all the magnetic moments of the orbital motions and spins of all the electrons in the atom. Each atom of a ferromagnetic material has an appreciable net magnetic moment, and, in addition, all the magnetic moments of the atoms within a domain are in alignment in the same direction, so the domain is magnetized to saturation. Thus, each domain has a sizable net magnetic moment. The magnetic dipole moment of a bar magnet is the vector *sum of the dipole* moments of all the *domains in* the magnet (see DIPOLE MOMENT). The preceding discussion gives the *fundamental reason* why a magnet produces a net magnetic field external to itself, and why it can experience a net torque when properly placed in a magnetic field. See ELECTRON SPIN.

North-south orientation. If a bar magnet is free to turn in the earth's magnetic field, it will set itself in a generally north-south direction. The north-seeking end is called the north pole of the magnet and the other end, the south pole. See GEOMAGNETISM.

Magnetic dipole moment. Consider the torque experienced by a bar magnet in any uniform magnetic field of known magnetic flux density B and, as shown in Fig. 1, with its axis at an angle β with respect to the direction of the applied field. The results of a systematic experiment show that (1) the torque τ tries to align the axis of the magnet parallel with the magnetic flux lines of the field; (2) τ is independent of the axis of rotation selected, therefore τ is due to a couple (equivalent to the action of two equal and opposite forces); (3) τ is directly proportional to B; and (4) τ is directly proportional to sin β. Hence, using M as a constant of proportionality,

$$\tau = MB \sin \beta \qquad (2)$$

The value of M is found to be constant for a given magnet, but different for different magnets, so M belongs to the magnet and is called the magnetic dipole moment of the magnet. *If $\beta = 90°$, $M = \tau/B$*, and the magnetic dipole moment of a magnet *is defined as equal to the moment of force that the* magnet experiences when it is perpendicular to a *magnetic field of unit magnetic flux density. In the* mks system, as Eq. (2) shows, the units of M are *newton-meter^2/weber or, the equivalent, ampere-*meter^2.

The definition of M given by this empirical method is the same as the one given earlier, which *says that M is the vector sum of the magnetic mo*ments of the domains of which the magnet is composed. This follows from the fact that the magnet experiences a torque in Fig. 1 only because the applied B field exerts a torque on the individual

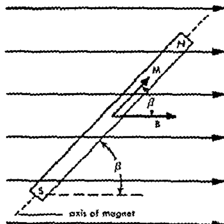


Fig. 1. Permanent bar magnet located in a uniform magnetic field.

atoms of the bar as a result of the spins and orbital motions of their electrons. It is known that electron spin is much more important than electron orbital motion in the magnetic behavior of the atoms of ferromagnetic materials.

Pole strength. Figure 2 is a photograph of an iron-filing map of the magnetic field of a permanent bar magnet. The iron filings arrange themselves like tiny compass needles along the magnetic flux lines. The regions called the poles are visible because the iron filings cluster in these regions and stand on end directly over the poles. The poles are simply regions where magnetic flux lines from the internal domains enter and leave the magnet. (The lines leave at the north pole and enter at the south pole.) However, one may look upon these poles as fictitious seats of magnetic effects and assign a pole strength m to each of them. When this pole point of view is used, the fictitious poles replace the uncompensated electron spins and orbital motions as the source of the magnetic behavior. Many complicated problems are considerably simplified by the use of the pole approach to behavior of magnets.

From the iron-filing map, one can more or less locate the centers of the poles. The effective magnetic length l of the magnet is the distance between the poles, as shown in Fig. 3. With a long slender magnet, l can be determined with fair accuracy. The pole strength m of each pole is defined as the quotient obtained when the magnitude M of the magnetic moment of the magnet is divided by the magnetic length l, or

$$m = M/l \qquad (3)$$

The units of m in the mks system, using the Sommerfeld proposal, are ampere-meters. (By the mks Kennelly proposal $M = \mu_0\tau/B$ when $\beta = 90°$, and M has the units of weber-meter; $m = M/l$ and thus the unit of M is the weber.)

[illegible]

two forces are to be of such magnitude that they

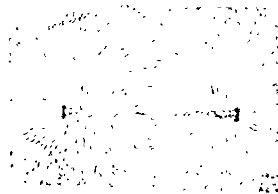


Fig 2. Photograph of the iron-filing map of the magnetic field of a permanent bar magnet. Note that the magnetic flux lines can be traced by the lines of iron filings, which act like tiny compass needles.

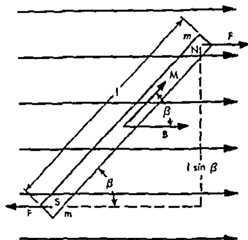


Fig. 3. Forces F on magnet poles in a magnetic field.

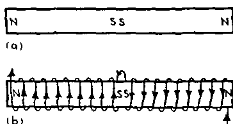


Fig. 4. (a) Magnet with consequent poles. (b) Diagram to show windings of coils which could produce the polarity shown in a.

will produce the actual torque experienced by the magnet. The two forces must be equal and opposite because, as already pointed out, the torque on a magnet in a uniform magnetic field is produced by a couple, and these two equal forces constitute such a couple. From Fig. 3, $\tau = Fl \sin \beta$, and from Eqs. (2) and (3), $\tau = mlB \sin \beta$; thus

$$F = mB \qquad (4)$$

is the force on each magnet pole in a magnetic field of flux density B. It follows from Eq. (4), and the fact that F is the same on both poles, that the

[illegible]

a magnet beyond the normal two. Figure 4 shows a magnet with a north pole at each end and the corresponding south poles together in the center and a diagram of the coil winding which could produce this type of polarity. *See* Electret; Electromagnet; Magnetic field; Magnetic flux; Magnetic materials; Magnetism; Magnetization; Magnetostatics. [R.P.WI.]

Bibliography: R. P. Winch, *Electricity and Magnetism*, 1955.

Magnet wire

The insulated copper wire used in the coils of electromagnets and other electromagnetic devices. It is single-strand wire insulated with paper, cotton,

enamel, varnish, glass, or combinations of these. A good insulation for magnet wire has high dielectric strength and occupies little space, since coil size is usually kept as small as possible.

Almost all magnet wire now in use is insulated, soft-drawn, electrolytic copper wire. Occasionally small amounts of silver (approximately 30 oz/ton) are added to improve the high-temperature tensile strength. For extremely high temperatures, anodized aluminum wire should be considered.

The choice of insulation also depends on temperature. The American Institute of Electrical Engineers (AIEE) Standards (1947) set five classes of insulation, based on the nature of the material and its high-temperature limitations. These are summarized in the following list.

Class	*Temperature limit*	*Acceptable materials*
O	90°C	Cotton, paper, silk, and similar organic materials
A	105°C	Materials of Class O when impregnated or immersed in a liquid dielectric; also phenolic resins, cellulose acetate, enamel and varnish
B	130°C	Mica, asbestos, fiberglass, and similar inorganic compounds with organic binding substances
H	180°C	Mica, asbestos, fiberglass, and similar inorganic compounds with silicone binding substances
C	no limit selected	Mica, porcelain, glass, quartz, and similar inorganic compounds

Paper-covered wire has a better space factor (ratio of copper cross section to total cross section) than cotton-covered and is superior in dielectric strength. It is widely used in transformer windings. The use of plain cotton-covered wire is comparatively rare. The cotton makes no contribution to the dielectric strength, but acts only as a spacer with air as the actual dielectric. More common applications for cotton are those in which the cotton-covered winding is impregnated with phenolic resin or other potting compound after the coil has been formed, or where the cotton is placed over an oleoresinous enamel coating during manufacture to protect it during the winding process. For heavy wire, such as is used in rotating machinery, cotton-covered enamel or impregnated cotton is a practical insulation. The impregnated cotton develops considerable mechanical strength, and the cotton cover absorbs vibration which might otherwise cause insulation wear and result in short circuits.

When space is an important factor, coatings of vinyl acetal resin varnishes, such as Formvar, are practical. They have superior resistance to mechanical abuse and may be used for winding coils without the additional protection of a cotton covering. Formvar is superior to enamel in its film flexibility, toughness, and ability to cling to the conductor when it is stretched. It is used in machinery applications where cotton-covered windings formerly were required. It is not soluble in conventional thinners. For American wire gage (AWG) no. 25 wire, the space factor for plain enamel-covered wire is 0.89; for Formvar, 0.86; for single cotton-covered, 0.5; and for double cotton-covered, 0.31. Both Formvar and plain enamel wire are available in multiple coats where higher dielectric strength is required and are considered Class A insulation.

Plain enamel is used in automobile ignition coils, magneto coils, small relays, fluorescent light ballasts, oil-burner transformers, and in common electronic applications. Its dielectric strength is greater than 1000 volts per millimeter thickness.

Glass-fiber-covered wire is used occasionally where higher temperatures are to be met (Class B). At higher (Class H) temperatures the glass fibers are bonded to the wire with a silicone varnish. Not only is the temperature rating improved (200°C), but its insulating quality and resistance to humidity are increased.

At even higher temperatures (500°C) anodized aluminum wire has been successfully used. Copper-nickel wire coated with anodized aluminum is reportedly operable to temperatures as high as 2000°C. *See* CONDUCTOR, ELECTRIC; INSULATION, ELECTRIC. [J.E.C.]

Bibliography: Anaconda Copper Company General Catalog; A. E. Knowlton (ed.), *Standard Handbook for Electrical Engineers*, 9th ed., 1957; H. Pender and W. A. Del Mar, *Electrical Engineers' Handbook*, 4th ed., 1949.

Magnetic amplifier

A device that employs saturable reactors to modulate the flow of alternating-current electric power to a load in response to a lower energy level direct-current input signal. Magnetic amplifiers are often referred to as power amplifiers, because they are well suited for use in driving electric motors and other output devices. See SERVOMECHANISM.

The concept of a saturable reactor is illustrated by the system shown in Fig. 1. The saturable reactor is much like an ordinary transformer, but it is operated with currents in its windings that can readily saturate the core material. The core is made of one of several special magnetic materials, rather than transformer iron, to achieve a given de-

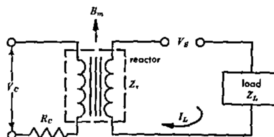


Fig. 1. saturable reactor circuit.

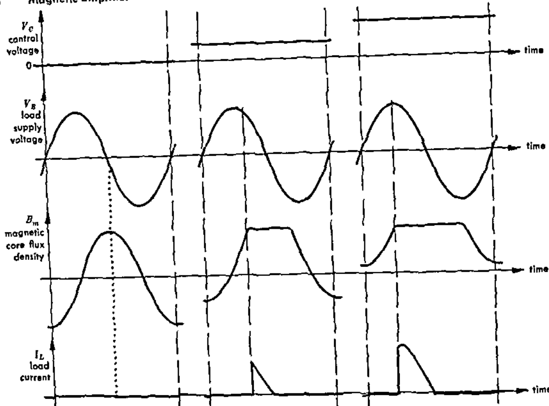


Fig 2. Voltage, flux, and current waveshapes for a single-core saturable reactor circuit with resistive load.

sired saturation characteristic. One of the most common core materials is Delta Max, which has a sharp saturation characteristic, high permeability, and a nearly rectangular flux density vs. magnetizing force characteristic. When the dc control-signal voltage V_C is zero in the system of Fig. 1, the saturable reactor is not quite saturated and it acts as a large inductance in the load circuit, thereby generating a back electromotive force which is nearly equal to and 180° out of phase with the voltage of the ac supply V_S. In other words, the impedance Z_r of the unsaturated reactor is much larger than the impedance Z_L of the load, and acts to block the flow of current from the ac supply through the load circuit. When the dc control-signal voltage is increased, the reactor (which was close to saturation with zero dc control signal) is caused to saturate during the part of each cycle when the magnetizing currents of the control and load windings are acting together. The duration of the period of saturation increases with increasing dc control voltage. Similarly, if the dc control signal had decreased from a zero value to a negative value, the reactor would have started to saturate during the other half cycle of the ac sinusoid. As soon as the reactor saturates, the impedance of the load winding suddenly drops to a small fraction of its unsaturated value (its back emf falls to a low value). Full ac supply voltage then acts to drive the load, and a pulse of current flows through the load until the reactor is desaturated a short time later as the load current from the ac supply decreases to zero. The shape of the load-current pulse depends on the exact nature of the load impedance. Figure 2 is a qualitative illustration of the load current at various dc control voltages for a purely resistive load.

The saturable reactor is essentially a synchronous switch which closes for a controlled portion of the ac supply cycle. The duration, and therefore the energy content, of the load-current pulse is a function of the magnitude of the dc control voltage, increasing with increasing dc control voltage.

Magnetic amplifiers often employ two saturable reactors combined to drive a single load from a single dc control signal. The system shown in Fig. 3, a series-connected arrangement employing two saturable reactors, overcomes many of the limitations of the single-reactor system shown in Fig. 1. The series arrangement makes it possible to obtain a controlled pulse during each half-cycle of the ac power sinusoid, and the problem of minimizing the effects of pulses in the control circuit is considerably simplified. Because of the pulses of current induced in the control windings, special care must be exercised in the design of the amplifier (or other device) that provides the dc control signal for a magnetic amplifier. Magnetic amplifiers are very nonlinear in their operation, and detailed analysis of their characteristics requires careful application of the theory of nonlinear magnetic circuits.

The speed of response of magnetic amplifiers is often severely limited by the frequency of the ac power source, because the speed depends on the

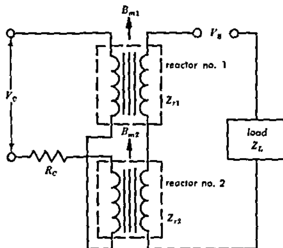


Fig. 3. Series-connected saturable reactors.

period of the ac power source. Impedance-matching problems at the input can also make it difficult to provide desired speed of response from the amplifier which drives the magnetic amplifier input. *See* SATURABLE REACTOR. For other types of amplifiers *see* AMPLIFIER. [J.L.SH.]

Magnetic circuits

The analysis of the relationship of the magnetomotive force (mmf), the magnetic flux, and the reluctance of the magnetic path. The relationship of these quantities takes the same form as Ohm's law for electric circuits, which relates electric current voltage and resistance. For magnetic circuits, the relationship is

$$R = \text{mmf}/\Phi \qquad (1)$$

where R is the reluctance and Φ the flux. The permeability μ is related to R by

$$R = L/\mu A \qquad (2)$$

where L is the length and A is the cross-sectional area of the magnetic path. The mmf is often expressed in ampere-turns NI, the product of the number of turns N and the current I needed to produce the field.

Per-unit quantities of flux density B and magnetomotive force H are usually more useful than total quantities.

$$B = \Phi/A \qquad H = NI/L \qquad (3)$$

Circuit calculations. The method of calculating magnetic circuits consists in computing the relation between the desired flux density and the required magnetizing force in each separate part of the circuit and then combining the results. This is necessary because of the nonlinear nature of the problem. Two different methods are required for finding this relationship, depending upon whether the path is magnetic or nonmagnetic.

Air-gap calculations. The most common nonmagnetic portion of magnetic circuits is the air gap. An air gap is unavoidable in most electromagnetic devices, and since the reluctance of air may be 10,000 times as great as that of iron or steel, an air gap contributes a major effect to magnetic circuits of which it is a part. Since saturation does not occur in air, a linear relation may be written

$$NI = 0.313\,BL \qquad (4)$$

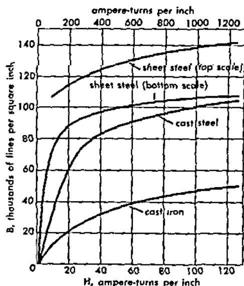
where the constant 0.313 is the reciprocal of the permeability in the English (inch) system of measurements.

As an example, if it is desired to find the number of ampere-turns necessary to produce a total flux of 10^6 lines across an air gap 0.5 in. long and 25 in.^{2} in cross section

$$NI = 0.313\left(\frac{10^6}{25}\right)0.5 = 6260 \text{ amp-turns}$$

Magnetic material calculations The calculations for magnetic materials must be accomplished differently because a linear relationship, such as Eq. 4, is destroyed by saturation. For magnetic materials a magnetization curve (or B-H curve) of the material must be used (see illustration). The steps in the calculations are (1) divide total flux by the cross-sectional area to obtain the flux density, (2) refer to B-H curve for the particular material employed, to determine H for the particular B used, and (3) multiply H by L to find the NI required. This procedure must be repeated for each change in cross-sectional area and for each piece of different magnetic material.

If the configuration is a series circuit, all of the values of ampere turns found for each of the segments, including air gaps, are added to obtain the total NI required to establish the assumed flux. To illustrate this let it be assumed that the air gap



Magnetization (B-H) curves (From A. Gray and G. A. Wallace, Principles and Practice of Electrical Engineering, 7th ed., 1955)

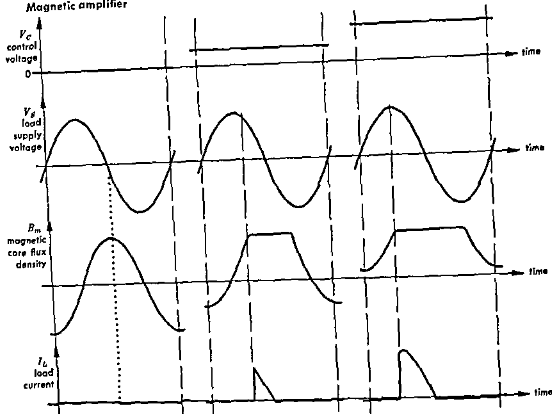


Fig 2. Voltage, flux, and current waveshapes for a single-core saturable reactor circuit with resistive load.

sired saturation characteristic. One of the most common core materials is Delta Max, which has a sharp saturation characteristic, high permeability, and a nearly rectangular flux density vs. magnetizing force characteristic. When the dc control-signal voltage V_C is zero in the system of Fig. 1, the saturable reactor is not quite saturated and it acts as a large inductance in the load circuit, thereby generating a back electromotive force which is nearly equal to and 180° out of phase with the voltage of the ac supply V_g. In other words, the impedance Z_r of the unsaturated reactor is much larger than the impedance Z_L of the load, and acts to block the flow of current from the ac supply through the load circuit. When the dc control-signal voltage is increased, the reactor (which was close to saturation with zero dc control signal) is caused to saturate during the part of each cycle when the magnetizing currents of the control and load windings are acting together. The duration of the period of saturation increases with increasing dc control voltage. Similarly, if the dc control signal had decreased from a zero value to a negative value, the reactor would have started to saturate during the other half cycle of the ac sinusoid. As soon as the reactor saturates, the impedance of the load winding suddenly drops to a small fraction of its unsaturated value (its back emf falls to a low value). Full ac supply voltage then acts to drive the load, and a pulse of current flows through the load until the reactor is desaturated a short time later as the load current from the ac supply decreases to zero. The shape of the load-current pulse depends on the exact nature of the load impedance. Figure 2 is a qualitative illustration of the load current at various dc control voltages for a purely resistive load.

The saturable reactor is essentially a synchronous switch which closes for a controlled portion of the ac supply cycle. The duration, and therefore the energy content, of the load-current pulse is a function of the magnitude of the dc control voltage, increasing with increasing dc control voltage.

Magnetic amplifiers often employ two saturable reactors combined to drive a single load from a single dc control signal. The system shown in Fig. 3, a series-connected arrangement employing two saturable reactors, overcomes many of the limitations of the single-reactor system shown in Fig. 1. The series arrangement makes it possible to obtain a controlled pulse during each half-cycle of the ac power sinusoid, and the problem of minimizing the effects of pulses in the control circuit is considerably simplified. Because of the pulses of current induced in the control windings, special care must be exercised in the design of the amplifier (or other device) that provides the dc control signal for a magnetic amplifier. Magnetic amplifiers are very nonlinear in their operation, and detailed analysis of their characteristics requires careful application of the theory of nonlinear magnetic circuits.

The speed of response of magnetic amplifiers is often severely limited by the frequency of the ac power source, because the speed depends on the

not upon the properties of the surrounding medium. This definition of H is only partial because it does not include contributions by magnetic poles if they are present in the neighborhood.

The mks unit of magnetic field strength appears from the defining equation, when current is in amperes and dl and r are in meters, as ampere per meter. Because many equations derived from the defining equation involve the number of turns N of a coil times the current, the ampere-turn per meter is also used as an equivalent unit.

Magnetic poles. A body can be magnetized by bringing it into a magnetic field due to currents or magnets (*see* MAGNETIZATION). Except in the case of a ring magnetized along its circumference, the field associated with a magnetized body extends to the region surrounding the body. The external effect usually appears in limited regions of the body called poles. A magnetized bar of iron has two poles, one at either end; and from the fact that the bar will set itself in an approximate north-south direction in the earth's field, it appears that there are two kinds of poles. The pole that is at the north end of the bar is called a north-seeking pole; that at the south end is called a south-seeking pole. The two poles at the ends are merely indications of the continuous magnetization within the body. An indication of the validity of this statement is the fact that when a bar magnet is broken into two parts, two new poles appear at the break, and the orientation of the poles in each fragment is the same as it was in the original magnet.

It is observed that magnetic poles exert forces on each other and upon moving charges in the region near the poles. There is a field near the pole, and the pole may be considered as the cause of that field.

If the poles of a magnetized body are small enough that they may be considered point poles, the force that one pole exerts upon another is found to be proportional to the product of the pole strengths m and m' and inversely proportional to the square of the distance r between them. This statement is called Coulomb's law of magnetostatics.

$$F = k' \frac{mm'}{r^2}$$

The proportionality factor k' depends upon the units used and upon the medium between the poles. For empty space in the mks system, k' is assigned the value $1/4\pi\mu_0$. The unit of pole strength associated with this choice is the weber. *See* WEBER; *see also* MAGNET; MAGNETOSTATICS; PERMEABILITY, MAGNETIC.

The magnetic field strength or magnetic intensity H may be expressed as the force per unit north-seeking magnetic pole.

$$H = \frac{F}{m'}$$

Then, from Coulomb's law, the contribution of a point pole of strength m to the magnetic field strength near the pole is

$$H = \frac{F}{m'} = \frac{1}{4\pi\mu_0}\frac{m}{r^2}$$

The direction of H is away from north-seeking poles and toward south-seeking poles. The contribution of several poles is the vector sum of the contributions of the individual poles

$$H = \frac{1}{4\pi\mu_0}\sum\frac{m}{r^2} \quad \text{(vector sum)}$$

If the H due to poles is to be in the same units as the H due to currents, the unit of pole strength must be chosen properly. If H is to be in amperes per meter when μ_0 is in webers per ampere-meter and r is in meters, then m must be in webers.

If the poles are distributed over surfaces or throughout volumes, the summation becomes an integral:

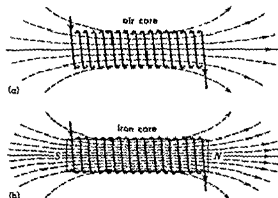
$$H = \frac{1}{4\pi\mu_0}\int\frac{dm}{r^2} \quad \text{(vector sum)}$$

The general expression for the field strength due to both currents and poles is given by

$$H = \frac{1}{4\pi}\int\frac{Idl\sin\theta}{r^2} + \frac{1}{4\pi\mu_0}\int\frac{dm}{r^2}$$

where the integrals represent vector sums.

If a toroidal coil has an iron core, the iron is magnetized by the current of the coil. The magnetic field strength within the core is given entirely by the first term of the equation for H, since there are no poles. For a long straight solenoid with an air core (part *a* of the illustration), H is entirely due to the current and is found by integration to be $H_I = NI/l$, where N is the number of turns of the coil, and l is the length of the solenoid. When an iron core is inserted into the solenoid (part *b*), the iron becomes magnetized, and poles appear at each end. The contribution of the current to H remains the same as before, but the second term now contributes to H components that are opposite in direction to H_I and that vary along the bar because



Flux in a solenoid (a) containing only air (essentially vacuum), and (b) containing an iron core.

of the variation in distances from the poles and because of the variation of the permeability of the iron. The effect of the poles on H is greatest at the ends near the poles. The magnetization of the iron is not uniform. The effect of the poles is essentially a demagnetizing effect; it is large for short magnets, and becomes negligible at the center of a long magnet.

Lines of force. As in the case of magnetic induction B, which can be represented by lines called magnetic flux, the vector quantity H may be represented by lines called lines of force. The number of lines of force per unit area of a surface perpendicular to H is made equal to the value of H. The direction of the lines of force is the direction of the field The lines of force are closed curves, as are the lines of induction. *See* LINES OF FORCE; MAGNETIC FLUX

Energy of a magnetic field. Consider a Rowland

found from Ampère's law to be

$$H = \frac{NI}{l}$$

where l is the mean circumference of the ring.

In building up the current in the coil, energy must be supplied that becomes energy of the magnetic field. This energy is

$$W = \tfrac{1}{2}LI^2$$

where L is the self-inductance of the coil (*see* INDUCTANCE). But

$$L = \frac{N\Phi}{I}$$

where Φ is the flux, and

$$W = \frac{1}{2}\frac{N\Phi}{I}I^2 = \tfrac{1}{2}N\Phi I = \tfrac{1}{2}NBAI$$

$$= \frac{1}{2}\frac{NI}{l}BV = \tfrac{1}{2}HBV$$

where V is the volume of the core, and A is the mean cross-sectional area of the ring. The energy per unit volume of the field is

$$\frac{W}{V} = \tfrac{1}{2}HB$$

If μ is the permeability of the core, $B = \mu H$ and the energy density of the field may be written

$$\frac{W}{V} = \tfrac{1}{2}HB = \tfrac{1}{2}\mu H^2 = \frac{1}{2}\frac{B^2}{\mu}$$

Magnetic potential. When a magnetic pole is in a magnetic field, there will be a force F acting on it. If it is moved a distance ds in the field, work dW is done against the field given by

$$dW = -F\cos\theta\, ds = Hm\cos\theta\, ds$$

where θ is the angle between the positive sense of H and the positive direction of s. The magnetic potential difference V may be defined as the work done per unit pole in taking the pole from one point to the other:

$$dV = \frac{dW}{m} = -H\cos\theta\, ds$$

or

$$V = -\int_{s_1}^{s_2} H\cos\theta\, ds$$

The resulting equation for the magnetic potential does not include the concept of the magnetic pole, and can be used as a defining equation for the magnetic potential. The integral represents the sum along a path of the products of ds and the component of H in the direction of ds. Such an integral is called a line integral and is represented by the symbol $\oint$. The defining equation for magnetic potential may be written

$$V = -\oint H\cos\theta\, ds$$

From this equation the relationship between magnetic potential and H may be written as

$$H\cos\theta = -\frac{dV}{ds}$$

Thus the component of field strength in any direction is the negative of the magnetic potential gradient in that direction. This statement is similar to the relation between electric field intensity and electric potential gradient.

The rise in magnetic potential around any closed path in a magnetic field may be deduced from consideration of the field about a long straight conductor (*see* BIOT-SAVART LAW). For this special case, the lines of force are concentric circles about the conductor and at a distance a from the conductor,

$$H = \frac{I}{2\pi a}$$

For a path that follows a circular line of force in a sense opposite to that of H, $\theta = 180°$ and $\cos\theta = -1$. Then the rise in magnetic potential around the closed path is

$$V = -\oint H\cos\theta\, ds = \int_0^{2\pi a}\frac{I}{2\pi a}\,ds = I$$

This result is independent of the radius of the circle followed and, in fact, is independent of the shape of the path, since any path may be resolved into components that are along circles and along radii. The contributions of the radial parts are zero, since they are perpendicular to H.

The result deduced from this special case may be generalized for a closed loop of any shape. The *rise in magnetic potential around the path is equal to the line integral of* $H\cos\theta$ *around the path, and this in turn is equal to the current through the surface bounded by the path*

$$\oint H\cos\theta\, ds = I$$

If the path taken includes no current, the integral is zero. If there are N equal currents inside the path, the integral becomes NI.

By analogy to an electric circuit, in which the line integral of the electric intensity around the circuit is the electromotive force of the circuit, the rise in magnetic potential around the closed path, that is, the line integral of $H \cos \theta$ around the path, is called the magnetomotive force (mmf). Thus, the mmf of a coil of N turns in which there is a current I is NI. *See* MAGNETOMOTIVE FORCE; *see also* ELECTRIC FIELD. [K.V.M.]

Bibliography: S. S. Attwood, *Electric and Magnetic Fields*, 3d ed., 1949; W. B. Boast, *Principles of Electric and Magnetic Fields*, 2d ed., 1956; L. Page and N. I. Adams, *Principles of Electricity*, 3d ed., 1958; F. W. Sears, *Principles of Physics*, vol. 2, 1951; R. P. Winch, *Electricity and Magnetism*, 1955.

Magnetic flux

Lines used to represent the magnetic induction **B** in a magnetic field. *See* INDUCTION, MAGNETIC; MAGNETIC FIELD.

The vector quantity **B** is defined from the equation of force **F** on a charge q moving with speed v in a magnetic field at an angle θ with the direction of the field. The magnitude of **B** is

$$B = \frac{F}{qv \sin \theta}$$

Lines of flux used to represent the field in magnitude and direction are selected so that they are parallel to **B** at each point, and the number of lines of flux per unit area of a surface perpendicular to the field is equal to B. The total number of lines of induction through a surface is the magnetic flux Φ. Magnetic induction is the flux per unit area or flux density. The flux $d\Phi$ through an element of area $d\mathbf{A}$ perpendicular to **B** is

$$d\Phi = \mathbf{B}\, d\mathbf{A}$$

If the surface is not perpendicular to **B**

$$d\Phi = B \cos \theta \, dA$$

where θ is the angle between **B** and the normal to the surface. Then $B \cos \theta$ is the normal component of **B**. For a surface of any orientation and for varying **B**, the flux through the surface is

$$\Phi = \int B \cos \theta \, dA$$

where the integral is taken over the whole surface.

Since lines of induction are always closed curves, it follows that for any closed surface in the magnetic field, every line that enters the closed surface must leave it. Therefore, the integral of the normal component of **B** over any closed surface in the magnetic field must be zero.

$$\int B \cos \theta \, dA = 0$$

In the mks system, a line of induction is called a weber, and the magnetic induction or flux density is in webers per square meter. Since, from the defining equation for magnetic induction, the unit of **B** is the newton/ampere-meter, that unit is equivalent to the weber per square meter. *See* FLUXMETER. [K.V.M.]

Magnetic lens

A magnetic field with axial symmetry capable of converging beams of charged particles of uniform velocity and of forming images of objects placed in the path of such beams. Magnetic lenses are employed as condensers, objectives, and projection lenses in magnetic electron microscopes, as final focusing lenses in the electron guns of cathode-ray tubes, and for the selection of groups of charged particles of specific velocity in velocity spectrographs.

Magnetic lenses may be formed by solenoids or helical coils of wire traversed by electric current, by axially symmetric pole pieces excited by a coil encased in a high-permeability material such as soft iron, or by similar pole pieces excited by permanent magnets. In the last two instances the armatures and pole pieces serve to concentrate the magnetic field in a narrow region about the axis.

Magnetic lenses are always converging lenses. Their action differs from that of electrostatic lenses and glass lenses in that they produce a rotation of the image in addition to the focusing action. For the simple uniform magnetic field within a long solenoid the image rotation is exactly 180 degrees. Thus a uniform magnetic field forms an erect real image of an object on its axis. This image has unity magnification and is at a distance from the object equal to

$$(8\pi^2 m\phi/eB^2)^{1/2} = 21.08\phi^{1/2}/B$$

Here m is the mass of the particles, e their charge, ϕ the accelerating potential of the particles, and B the magnetic induction on an axis of symmetry of the field. The numerical coefficient 21.08 applies for electrons, with ϕ in volts and B in gauss.

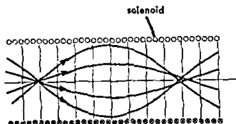
For short magnetic lenses, or lens fields which are short compared to the focal length, both the magnification and image position depend on the position of the object. The focal length f is given by

$$\frac{1}{f} = \frac{e}{8m\phi} \int B^2 \, dz = \frac{0.022}{\phi} \int B^2 \, dz \quad \text{cm}^{-1}$$

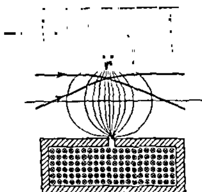
The integration is carried out over the extent of the lens field along the axis of symmetry (the z axis). The numerical coefficient 0.022 again applies for electrons. At the same time the field produces an image rotation through an angle

$$\theta = \left(\frac{e}{8m\phi}\right)^{1/2} \int B \, dz = \frac{0.147}{\phi^{1/2}} \int B \, dz$$

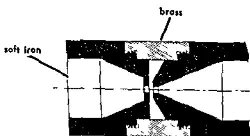
Thus the magnetic lens field formed by two identical coils in tandem, [illegible] oppositely di-rected currents, is rotat[illegible] a spec[illegible] ample of a short lens, a [illegible] loop



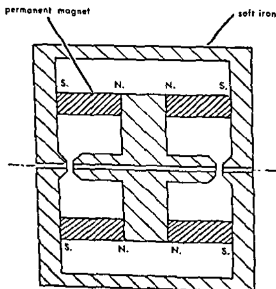
(a)



(b)



(c)



(d)

Magnetic lenses. (a) Uniform magnetic field. (b) Short magnetic lens formed at gap in soft-iron casing about coil. (c) Pole pieces for magnetic electron microscope objective. (d) Double magnetic lens excited by permanent magnets. (*From E. G. Ramberg and G. A. Morton, J. Appl. Phys., vol. 10, 1939, and J. Hillier and E. G. Ramberg, J. Appl. Phys., vol. 18, 1947*)

of radius r traversed by current I produces a lens with focal length

$$f = 96.8r\phi/I^2 \quad \text{cm}$$

Here ϕ is in volts and I in amperes. The image rotation for this lens is

$$\theta = 0.185I/\phi^{1/2} \quad \text{radian}$$

[E.G.RA.]

Bibliography: *See* ELECTRON MOTION IN VACUUM.

Magnetic materials

Although all materials have magnetic properties of some kind (diamagnetic, paramagnetic, or ferromagnetic), the term magnetic materials is customarily applied only to substances which exhibit ferromagnetism. *See* FERROMAGNETISM.

Such materials are of two kinds: (1) those easily magnetized, called soft magnetic materials, and suitable for such devices as transformers; and (2) those that retain their magnetism with great tenacity, designated hard magnetic materials, and suitable for fabrication into permanent magnets. For practical purposes the former are used in transformers, electric motors and generators, relays, telephone receivers, sonar equipment, radar, and other devices; the latter find application in loudspeakers and electrical measuring instruments. Several million tons of magnetic materials are manufactured annually.

The properties of soft magnetic materials that are important for their application in industry are normally low coercive force and high permeability. Extremely high permeability at high magnetic induction (B = 10,000–20,000 gauss) is required of materials used in power transformers and relays, and high permeability at low inductions (B = 10–1000 gauss) is necessary for materials used in communication. In permanent-magnet materials, high coercive force (H_c = 50–3,000 oersteds) and high remanence (B_r = 5,000–12,000 gauss are required. *See* INDUCTION, MAGNETIC; MAGNETIZATION; PERMEABILITY, MAGNETIC.

Soft materials. Unalloyed iron and iron to which a trace of silicon has been added comprise the bulk of the soft materials used for the production and distribution of electric power In the communications industry, many of the magnetic alloys are of the permalloy type and contain large percentages of nickel and iron and often moderate amounts of other elements. *See* IRON-SILICON ALLOY; PERMALLOY.

Since World War II, a new class of substances, nonmetallic ferrites, has found increasing use.

Table 1. Some properties of high-permeability materials

Material	Composition, % wt, remainder iron	μ_o, Initial permeability, gauss/oersted	μ_m, Maximum permeability, gauss/oersted	H_c, Coercive force, oersteds	B_s, Saturation induction, gauss	θ, Curie point, °C
Grain-oriented Fe-Si	3 Si	7,500*	55,000	0.1	20,000	740
Hot-rolled Fe-Si	4 Si	1,500*	7,000	0.3	19,500	730
Thermoperm	30 Ni				2,000	50
45 Permalloy	45 Ni	2,500	25,000	0.1	16,000	400
Hipernik	50 Ni	4,000	70,000	0.05	16,000	500
78 Permalloy	78.5 Ni	8,000	100,000	0.05	10,800	600
4–79 Permalloy	4 Mo, 79 Ni	20,000	100,000	0.05	8,700	460
Supermalloy	5 Mo, 79 Ni	100,000	1,000,000	0.002	7,900	400
Mumetal	5 Cu, 2 Cr, 77 Ni	20,000	100,000	0.05	6,500	
1040 alloy	3 Mo, 14 Cu, 72 Ni	40,000	100,000	0.02	6,000	290
Permendur	50 Co	800	5,000	2.0	24,500	980
Vanadium Permendur	1.8 V, 49 Co	800	4,500	2.0	24,000	980
Supermendur	2 V, 49 Co		60,000	0.3	24,000	980
45-25 Perminvar	45 Ni, 25 Co	400	2,000	1.2	15,500	715
Sendust	5 Al, 10 Si	30,000	120,000	0.05	10,000	500
Mn-Zn ferrite		1,500	2,500	0.2	3,400	130
Ni-Zn ferrite		2,500	5,000	0.1	3,200	140

* For $B = 100$, μ_o is lower and uncertain.

These materials are oxides having chemical formulas such as $NiFe_2O_4$ (nickel ferrite), $MnFe_2O_4$ (manganese ferrite), and combinations of these with the ferrites of zinc, magnesium, and other elements. The electrical resistivities of ferrites are 10^6–10^{12} times as high as those of magnetic alloys, and for this reason they are often much more suitable for use at the relatively higher radar and radio frequencies. Some of the important magnetic properties of commonly used soft magnetic materials are given in Table 1. For the composition of the ferrites listed in Table 1 *see* FERRITE.

Hard materials. Most of the magnetically hard materials used for permanent magnets are alloys

Table 2. Some properties of materials used for permanent magnets

Material	Composition, % by weight, remainder iron	H_c, coercive force, oersteds	B_r, remanence, gauss	$(BH)_m \times 10^{-6}$, gauss-oersteds	Preparation	Heat treatment	Mechanical properties
Carbon steel	0.9 C, 1 Mn	50	10,000	0.2	Hot roll, machine, punch	Q 800	Hard, strong
Tungsten steel	0.7 C, 0.3 Mn, 5 W	70	10,300	0.3	Hot roll, machine, punch	Q 850	Hard, strong
36 Co steel	0.7 C, 36 Co, 4 Cr, 5 W	240	10,000	1.0	Hot roll, machine, punch	Q 930	Hard, strong
Alnico 5	24 Co, 14 Ni, 8 Al, 3 Cu	600	12,500	5.0	Cast, grind	AF 1300 B 600	Hard, strong
Remalloy	12 Co, 17 Mo	250	10,500	1.1	Hot roll, machine, punch	Q 1200 B 700	Hard, malleable
Vicalloy	52 Co, 14 V	510	10,000	3.5	Cold roll, draw	B 600	Ductile
Cunife	20 Ni, 60 Cu	500	5,400	1.3	Draw, machine, punch	Q 1070 B 700 D,B 600	Ductile
Cunico	41 Co, 24 Ni, 35 Cu	660	3,400	0.8	Cold roll, machine, punch	Q 1080 B 625	Ductile
Pt-Cobalt	77 Pt, 23 Co	4,000	6,000	9.0	Cast	Q 1200 B 650	Hard, strong
Ferroxdur	100 $BaFe_{12}O_{19}$	1,600–2,000	2,000–3,000	0.9–3	Sinter	PF	Hard, brittle
Fine-powder Fe + Co	30 Co	1,000	9,000	5.0	Press	None	Brittle
Bismanol	100 MnBi	3,100	4,300	4.3	Sinter	PF	Brittle

Q, quenched from indicated temperature (°C) in oil or water; AF, cooled in magnetic field; PF, pressed in magnetic field; B, baked; D, drawn (wire).

of the Alnico type (*see* ALNICO). Properties comparable with those of the Alnicos can be obtained in fine pressed powders of iron or iron-cobalt alloy, or in those of certain nonmetallic compounds. Other types are used when ductility is required (*Cunife, Cunico, Vicalloy*). Information about some of the most used materials is given in Table 2. The "energy product" $(BH)_m$ given there is the principal figure of merit of a material for use in permanent magnets.

Other materials. In addition to the materials listed in the tables, there are many that have interesting properties and are of considerable scientific interest. These include the Heusler alloys such as Cu_2MnAl and Cu_2MnSn, Fe Co alloys (high magnetic saturation from 30 to 60% cobalt), Fe-Al alloys (4–15% aluminum), Perminvar alloys (Fe-Co-Ni), Ni-Mn and Zr-Zn alloys, some of the rare-earth metals and their alloys, and many ferrites.

Atomic structure. Ferromagnetic materials are composed essentially of atoms that are small permanent magnets; the magnetic moments of neighboring atoms are held in fixed orientation by means of the quantum mechanical forces of exchange. Most of the atoms which possess permanent moments are those of the transition elements (iron and rare-earth series); the moment is due to the imbalance of electron spins in one of the unfilled electron shells of the atom. The study of magnetic materials is useful in determining atomic structure and the nature and magnitude of interatomic forces in solids. *See* IRON ALLOYS [R.M BO.]

Bibliography: R. M. Bozorth, *Ferromagnetism*, 1951; D E. Gray, *American Institute of Physics Handbook*, 1957

Magnetic moment

The relationship between a magnetic field and the torque exerted on a magnet, a current loop, or a *charge that is moving in the field.*

When a magnet is placed in a magnetic field of strength H, there is a torque L exerted on the magnet by the field. The torque is a maximum when the axis of the magnet is perpendicular to the field. The ratio of the torque for this position to the strength of the field is called the magnetic moment M of the magnet.

$$M = \frac{L}{H}$$

See MAGNET

If a flat coil of wire of N turns and area A, in which there is a current I, is placed in a magnetic field of flux density B, the coil experiences a torque L given by

$$L = NIAB \sin \theta$$

where θ is the angle between the field and the normal to the plane of the coil. The torque is maximum when $\theta = 90°$, that is, when the plane of the coil is parallel to the field. The ratio of the maximum torque to the flux density B is the magnetic moment of the coil

$$M = \frac{L}{B} = NIA$$

Alternatively, the magnetic moment of the coil may be defined as the ratio of L to H. For this definition,

$$M = \frac{L}{H} = \mu_0 NIA$$

since $B = \mu_0 H$ in empty space, μ_0 being the permeability of empty space.

An electron in its orbit about an atomic nucleus has an orbital magnetic moment. If I represents the equivalent current as the electron moves in its orbit, and T is the time of one revolution

$$I = \frac{q}{T} = \frac{q\omega}{2\pi}$$

where ω is the angular velocity and q is the charge of the electron. Then the orbital magnetic moment is

$$M_o = IA = \frac{q\omega}{2\pi} A$$

where A is the orbital area.

If a charge is spinning, there is a charge in motion and thus an electric current. The spin is equivalent to a tiny current loop which has a magnetic moment. *See* ELECTRON SPIN; MAGNETIC MOMENT, ELECTRON.

Atomic nuclei also possess magnetic moments. For a discussion of these, *see* NUCLEAR MOMENTS. *See also* MAGNETON. [K.V.M.]

Bibliography: F. W. Sears, *Principles of Physics*, vol. 2, 1951; R. P. Winch, *Electricity and Magnetism*, 1955.

Magnetic moment, electron

The electron has magnetic properties by virtue of (1) its orbital motion about the nucleus of its parent atom and (2) its rotation about its own axis. The magnetic properties are best described through the magnetic dipole moment associated with (1) and (2). The classical analog of the orbital magnetic dipole moment is the dipole moment of a small current-carrying circuit. The electron spin magnetic dipole moment may be thought of as arising from the circulation of charge, that is, a current, about the electron axis; but a classical analog to this moment has much less meaning than that to the orbital magnetic dipole moment. The magnetic moments of the electrons in the atoms that comprise a solid give rise to the bulk magnetism of the solid. For a detailed discussion of magnetic moments of electrons, *see* ELECTRON SPIN; *see also* ATOMIC STRUCTURE AND SPECTRA; MAGNETISM. [P.KU.]

Magnetic recording

Process of recording information on paper- or plastic-base tape coated with magnetized ferromagnetic powder, or on magnetized wire. Although the

most widespread application of magnetic recording has been, and is, the recording of audio (sound) signals, other signals, such as information from data and computer systems, black-and-white and color television pictures, and radio signals from artificial satellites, can be recorded and subsequently played back.

The use of a moving magnetic wire as a means for reproducing audio signals was first demonstrated in 1898 (*see* WIRE RECORDING). However, it is only since the late 1940s that theories, techniques, and materials have been developed which make possible the reproduction of audio signals with a satisfactory performance from the standpoint of frequency range, signal-to-noise ratio, and distortion. Although satisfactory performance was obtained from wire, magnetic tape, consisting of a paper or plastic base coated with a thin layer of iron oxide, is now almost universally used in magnetic sound reproducers.

Monophonic system. A monophonic magnetic tape recording system consists of a magnetic head for producing a varying magnetic field, and a mechanism for moving the magnetic tape relative to the head, thereby recording magnetic signals in the magnetic tape corresponding to the electrical signals. A monophonic magnetic tape reproducing system consists of a magnetic head and a mechanism for moving the tape past the head by means of which the recorded magnetic variations are converted into electrical signals of approximately like form.

Recording system. The elements in a complete monophonic magnetic tape recording system are shown in Fig. 1. The first element is the acoustics of the studio or room. If more than one microphone is used (for example, when a soloist accompanies an orchestra, there is one microphone for the singer and one for the orchestra) the outputs of the two microphones may be adjusted for the proper balance by means of the mixers, devices which have two or more inputs and a common output. An electronic compressor is used to reduce a large amplitude range to that suitable for recording. A corrective electrical network called an equalizer provides the recording characteristic, which is described later. The attenuator provides a control on the over-all level fed to the power amplifier. The output of the power amplifier is fed to the recording head. The magnetic recording head, actuated by the amplifier, magnetizes the magnetic coating on the tape in a pattern which corresponds to the undulations in the original sound wave; that is, the magnetic flux pattern of the recorded tape consists of a series of magnetized sections. To overcome the nonlinearity of the tape or wire, a high-frequency signal, termed a bias, is fed to the recording head together with the audio signal. A monitoring system consisting of a complementary equalizer, attenuator or gain control, volume indicator, power amplifier, and loudspeaker is used to control the recording operation.

Recorders. A typical tape-transport recording and reproducing system is shown in Fig. 2. In the recording or reproducing process, the tape is transferred from the payoff reel to the takeup reel. The tape passes over three magnetic heads—the erase head, the record head, and the reproduce head. From the heads the tape passes between the capstan and a pressure roller. The capstan is driven at constant angular speed to insure constant linear speed of the tape as it passes over the heads. A braking force is applied to the payoff reel and a driving force to the takeup reel.

The standard tape speeds are 30, 15, 7½, 3¾, and 1⅞ in./sec. The higher speeds are used for high-quality professional recording; the lower speeds are used for commercial reproduction. The NARTB standard tape width is ¼ in. Single-, double- and quadruple-track recordings are used on

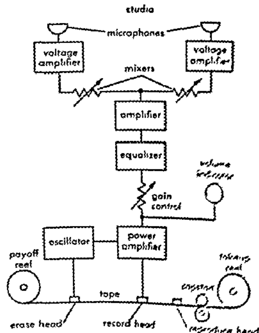


Fig. 1. Schematic arrangement of apparatus in a complete monophonic magnetic tape recording system

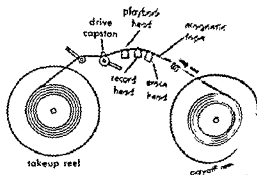


Fig. 2 Elements of a magnetic tape recording and reproducing system. (After H. F. Olson, Acoustical Engineering, Van Nostrand, 1957)

the ¼-in. tape. Wider tapes, higher and lower tape speeds, and a large number of tracks are used in some professional and special purpose magnetic tape machines.

Magnetic tape. The tape used in magnetic recording consists of a plastic base with a coating of magnetic oxide. The base material of Mylar or cellulose acetate varies in thickness from 0.0007 to 0.0015 in. The magnetic coating varies in thickness from 0.0004 to 0.0007 in The three standard thicknesses of magnetic tape are 2.2, 1.7, and 1.1 mils. A typical *B-H* characteristic of the magnetic coating is shown in Fig. 3, where *H* is the magnetizing force and *B* is the magnetic induction. The retentivity is of the order of 700 gauss, and the coercive force is of the order of 300 oersteds. *See* MAGNETIZATION.

Reels. The reel most common in mass use is 7 in. in diameter The 7-in. diameter reels hold 1200 ft of the standard 2.2-mil tape, 1800 ft of the long-playing 1.7-mil tape, and 2400 ft of the extra-long-playing 1.1-mil tape. Most professional machines use the NARTB standard 10.5-in. diameter reel. This reel holds twice as much tape as the 7-in. diameter reel.

Magnetic recording process. The recording process is depicted in Fig. 4. The passage of the tape past the recording head leaves a series of magnetized sections which correspond to the electrical signal applied to the head when the head gap was in contact with the tape at each of the sections.

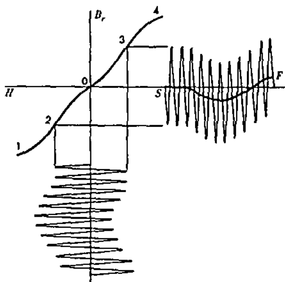


Fig. 5. Recording and reproducing characteristics for magnetic tape with a high-frequency bias. The characteristic 1, 2, 0, 3, 4 is the residual induction B_r produced by the magnetizing force *H* from the recording head. The high-frequency bias and audio signal are applied to the tape The resultant characteristic in reproduction is the characteristic *S-F*. (*After H. F. Olson, Acoustical Engineering, Van Nostrand, 1957*)

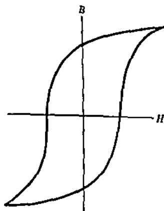


Fig. 3. Typical *B-H* magnetic characteristic of the iron oxide coating on magnetic tape

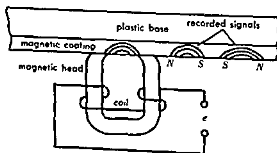


Fig. 4 Schematic diagram depicting the magnetic tape recording and reproducing process. (*After H. F. Olson, Acoustical Engineering, Van Nostrand, 1957*)

The magnetic material used in the coating of the plastic tape is of necessity nonlinear because it must possess retentivity to retain the magnetic signal applied to the tape in recording. The characteristic which depicts the magnetomotive force or magnetizing force *H* produced by the recording head in the magnetic tape and the residual induction B_r after the magnetic tape leaves the head is shown in Fig. 5. The nonlinear portion in the vicinity of 0 of Fig. 5 will produce distortion. Various means have been developed for reducing the effects of this nonlinear characteristic. The system which is universally used in sound reproduction is the one employing alternating current bias. The high-frequency signal, usually in the range of 50–150 kc/sec, is added to the audio signal in the recording head, there being no modulation of one signal by the other. The action of the high-frequency bias in reducing the effect of the nonlinear characteristic is shown in Fig. 5.

Magnetic heads. Sectional views of a single-channel magnetic recording or reproducing head are shown in Fig. 6. The magnetic head consists of a stack of laminations with an air gap where the head makes contact with the tape and another air gap in the rear portion of the head. Two coils surround the laminations. A schematic top view of the recorded magnetic track and sectional views of the magnetic head are shown in Fig. 6. In general, two single-channel tracks are recorded on the tape.

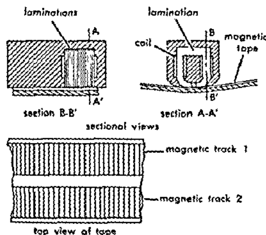


Fig. 6. Sectional views of a single channel magnetic head and the magnetic track on ¼-in. magnetic tape. In monophonic recording, magnetic track 1 is used for the forward direction and magnetic track 2 is used for the reverse direction. For stereophonic recording, a dual head and magnetic tracks 1 and 2 are used for the right and left channels.

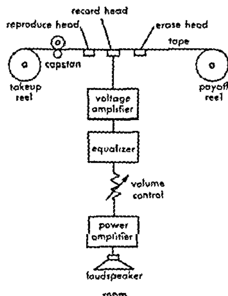


Fig 7. Schematic arrangement of apparatus in a complete monophonic magnetic tape reproducing system.

The two tracks are recorded in opposite directions, which eliminates rewinding in reproduction.

Reproducing system. The elements of a complete magnetic tape monophonic sound reproducing system are shown in Fig. 7. When the tape is reproduced, the magnetized sections move past the head and produce a magnetic field in the head which corresponds to the pattern of the tape. A voltage is developed in the coil of the head resulting from the variation in magnetic flux through the coil of the head. The electrical signal corresponds to the electrical signal applied to the recording head when the signal was recorded on the tape. The output of the head is amplified by a voltage amplifier, and the amplifier is followed by an equalizer. A tone-control system may also be employed. The output of the equalizer is fed to a volume control, the output of which is fed to a power amplifier. The output of the power amplifier drives a loudspeaker. The tape-transport mechanism is the same as that used for recording shown in Fig. 2.

Recording and reproducing characteristics In the recording and reproducing process there is a loss due to the finite gap length, that is, the distance between adjacent surfaces of the north-south pole pieces of a magnetic head In the recording process the gap length is of little importance because the recording process takes place from the trailing edge of the gap. This edge, rather than the gap, is of importance in the recording head. In the reproducing process the length of the gap determines the magnetomotive force. The loss due to the finite length of the gap in the reproducing head is given by

$$R = -20 \log \left[\frac{\sin (\pi d/\lambda)}{\pi d/\lambda}\right]$$

where R is the loss in decibels (db), d the length of the gap in centimeters, and λ the wavelength of the signal along the tape in centimeters.

The output of the reproducing head is

$$e = N\frac{d\phi}{dt}$$

where e is voltage output in abvolts, N the number of turns in the coil, ϕ the flux in the coil in gauss, and t the time.

If the amplitude of ϕ is constant, the voltage will increase at the rate of 6 db/octave. However, this characteristic must be multiplied by the gap loss. The open-circuit voltage response of a magnetic reproducing system is given by

$$R_M = 20 \log \left(\sin \frac{\pi d}{\lambda}\right)$$

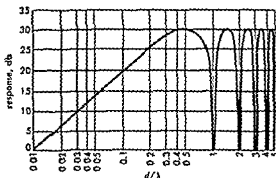


Fig. 8. The response of a magnetic tape reproducing system as a function of the ratio of gap length d to wavelength λ. (After H. F. Olson, Acoustical Engineering, Van Nostrand, 1957)

The open-circuit voltage frequency-response characteristic of a magnetic reproducing system is shown in Fig. 8. The zeros of response occur when the length of the gap is a multiple of a wavelength. The upper frequency limit of the recording range is the first zero of response. The ideal frequency-response characteristic of a magnetic reproducing system in terms of the frequency and gap is

$$R_M = 20 \log \sin \left(\frac{\pi df}{v}\right)$$

where f is frequency in cps and v is the tape speed in cm/sec

The NARTB standard reproducing characteristic for a tape speed of 15 in./sec is shown in Fig. 9. Some additional equalization is introduced in the low- and high-frequency ranges of the recording system to provide a uniform over-all frequency-response characteristic.

The signal-to-noise ratio for a high-grade magnetic tape reproducing machine should be at least 55 db.

Stereophonic system. A two-channel magnetic tape stereophonic sound reproducing system employing prerecorded magnetic tape was commercialized in 1956. The stereophonic magnetic tape player provides the reproduction of the original sound sources in auditory perspective; that is, the spatial relations of the original sound are substantially retained in the reproduction of the recorded sound.

Recording system. The elements of a complete stereophonic magnetic tape recording system are shown in Fig. 10 There are two identical channels of the type shown in Fig. 1 up to the magnetic recording head A two-channel magnetic head is shown in Fig. 11. The head is in reality two heads of the type shown in Fig. 6. The head shown in Fig. 11 is termed a stacked head because the gaps in the two heads are in a line The so-called staggered-head system, in which the two sections of the head are separated along the length of the tape, has been largely discontinued and the stacked head is now the standard arrangement Two track arrangements are used as shown in Fig. 11.

Reproducing system. The elements of a stereophonic magnetic tape reproducing system are shown in Fig. 12. There are two identical channels of the type shown in Fig. 7 following the head. The magnetic head is the same as that used for stereophonic recording.

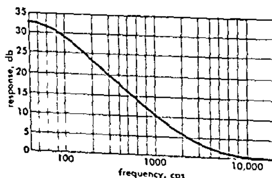


Fig 9. The NARTB standard magnetic tape reproducing characteristic.

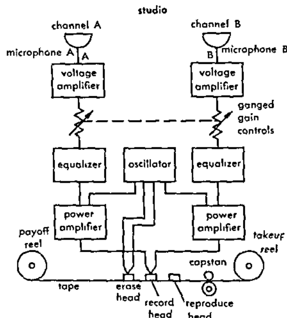


Fig. 10. Schematic arrangement of apparatus in a complete stereophonic magnetic tape recording system.

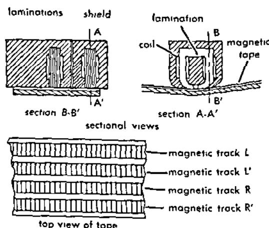


Fig 11. Sectional views of a two-channel stereophonic magnetic head and the quadruple magnetic tracks on ¼-in. magnetic tape. In the arrangement shown, tracks L and R are used with the tape moving in the forward direction, and the head is shifted and tracks L' and R' are used with the tape moving in the reverse direction.

Binaural system. The systems shown in Figs. 1 and 12 may be employed as a binaural sound reproducing system by mounting the two microphones in a dummy head and substituting two telephone receivers for the loudspeaker.

Binaural sound reproducing *systems have not* been commercialized on a wide scale because a set

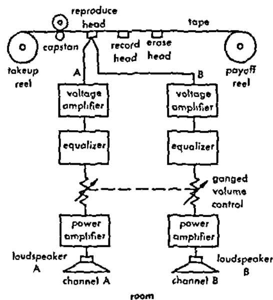


Fig. 12. Schematic arrangement of apparatus in a complete stereophonic magnetic tape reproducing system.

of earphones must be worn throughout the rendition of the program.

The advantage of the system is that the binaural effect is practically perfect with the result that the listener is in effect acoustically transferred to the point of sound pickup. For this reason binaural sound reproductions are particularly useful for subjective investigations.

Distortion and noise. The most common source of distortion in magnetic tape reproducing systems is due to the inherent nonlinear characteristics of the recording medium. This type of distortion can be reduced to a negligibly low level by proper design and operation of the ac bias system.

Another source of distortion, termed drop-out, is due to an absence of magnetic material on the tape with the result that there is a large reduction in output when this part of the tape passes the head.

Noise occurs in reproduction of magnetic tape due to the random distribution of the magnetic particles and nonuniform thickness of the magnetic coating.

Another source of distortion is due to the nonuniform motion of the magnetic tape past the recording and reproducing heads. It is manifested as a frequency modulation of the reproduced signal and is termed flutter and wow. *See* FLUTTER AND WOW.

Print-through is the transfer of magnetism from a magnetized layer to adjacent layers of tape. In this way a small amount of the original signal is produced in adjacent layers. This results in a pre- and post-echo when the tape is reproduced. Print-through is proportional to the time elapsed after recording and inversely proportional to the thickness of the base. The processing in the manufacture of the tape influences the amount of print-through. The level of the print-through is about 50 db below the level of the signal.

Telephone tape recording. A magnetic tape recording and reproducing system is used as an answering service when the telephone subscriber is absent. A suitable message can be recorded which will be reproduced when the telephone subscriber's number is called. In response to an incoming call, the machine answers with the prerecorded message, after which it switches to the recording condition and records the message which the caller wishes to deliver to the subscriber.

Television tape recording. Magnetic tape systems are used for recording and reproducing television signals in both monochrome and color. The video signal is recorded in one channel and the audio signal in another channel. A relatively high ratio of tape-to-head speed is required to record the video signal. For an extended discussion of video tape recording, *see* TELEVISION.

Data and computer recording. Magnetic tape recording and reproducing systems are used for the storage of information in all manner of data and computer systems. The information is stored on multiple magnetic tracks in the form of discrete pulses. A few hundred discrete pulses can be recorded per linear inch of the tape. The system is arranged so that any part of the recorded information can be retrieved at will. For an extended discussion of these storage techniques, *see* DATA PROCESSING SYSTEMS.

Prerecorded magnetic tapes. Prerecorded monophonic and stereophonic magnetic tapes containing program material similar to that of disk phonograph records have been commercialized on a wide scale. In the production of prerecorded magnetic tapes, the master tape is reproduced by a magnetic tape reproducer, and the output is fed to a bank of magnetic tape recorders so that many prerecorded magnetic tapes are produced at once.

Two types of prerecorded tapes have been commercialized on ¼-in. tape, namely, two tracks with a tape speed of 7½ in./sec; and four tracks, two in the forward direction and two in the reverse direction, with a tape speed of 3¾ in./sec. Magnetic-tape cartridges, in which the prerecorded magnetic-tape reels are enclosed in a case, have been commercialized. Magnetic-tape cartridge reproducers have also been commercialized. *See* OPTICAL RECORDING. [H.F.O.]

Bibliography: S. J. Begun, *Magnetic Recording*, 1955; J. Moir, *High Quality Sound Reproduction*, 1958; H. F. Olson, *Acoustical Engineering*, 1957; W. E. Steward, *Magnetic Recording Techniques*, 1958; J. Tall, *Techniques of Magnetic Recording*, 1958.

Magnetic relaxation

The relaxation or approach of a magnetic system to an equilibrium or steady-state condition as the magnetic field is changed. This relaxation is not instantaneous but requires time. The characteristic times are called relaxation times.

Magnetism is associated with angular momentum called spin, because it usually arises from spin of nuclei or electrons. The spins may interact with applied magnetic fields, the so-called Zeeman energy; with electric fields, usually atomic in origin; and with one another through magnetic dipole or [illegible]

which does not is called spin-spin relaxation. (As used here, the term lattice does not refer to an ordered crystal but rather signifies degrees of freedom other than spin orientation, for example, translational motion of molecules in a liquid.) The former is associated with the approach of the spin system to thermal equilibrium with the host material, the latter with an internal equilibrium of the spins among themselves. *See* SPIN (QUANTUM MECHANICS); see *also* MAGNETISM and the articles listed therein.

The measurement of relaxation times is of interest in the study of molecular motions and interactions. An important practical application of magnetic relaxation is the attainment of very low temperatures For example, a sample of a paramagnetic salt can be cooled by liquid helium to about 1°K, then subjected to a magnetic field. Heat, evolved in the magnetization, will be dissipated into the liquid helium If thermal contact to the helium is broken, and the magnetic field is then removed, the relaxation of the sample to more random orientation will be accompanied by a further decrease in temperature, in some cases to 0.01°K. A similar adiabatic demagnetization of copper, in this case involving nuclear spin, has cooled the sample to 0 00002°K. *See* ADIABATIC DEMAGNETIZATION.

Relaxation process. In an unmagnetized sample (ferromagnets are excluded for the moment) the spin orientations are random. In the presence of a static magnetic field and at thermal equilibrium, an excess of the spin vectors point in the lower [illegible]

turning process may go to either or both of two reservoirs. If the applied field is weak, the spin-spin energy constitutes an effective reservoir. If the applied field is strong, the spin-spin (free) energy is negligible and magnetization requires an interchange with the lattice. At low fields, the transfer of energy into the spin-spin reservoir may be thought of as corresponding to a heating of the spin system. The spin-lattice coupling will eventually bring the spins into thermal equilibrium with the lattice (by changing either the spin "temperature," the lattice temperature, or both). Thus for low fields the process of magnetization would exhibit two stages: (1) the redistribution of energy in the spin system, or spin-spin relaxation, and (2) equilibration of the spin system with the lattice, or spin-lattice relaxation. For high fields, only spin-lattice relaxation plays a role in establishing the magnetization.

Although relaxation exhibits itself in a macroscopic manner, an explanation of the rates of relaxation requires consideration of atomic processes. Two requirements must be met for spin-lattice relaxation: (1) there must be a coupling of the spin to the lattice, and (2) the coupling must be time-dependent, its frequency spectrum containing natural frequencies of the spin system (for example, the precession frequency of the spins in the applied field).

Types of relaxation. Relaxation has been studied for nuclear magnetism, electron paramagnetism, and ferromagnetism.

Nuclear relaxation. For nuclei, the Zeeman and spin-spin interactions are always important, and in some cases, the nonspherical charge distribution (electric quadrupole moment) of a nucleus affects the relaxation time. When the quadrupolar energy may be neglected, the magnetization usually approaches thermal equilibrium exponentially after application of a strong static field, the time constant being the spin-lattice relaxation time. Nuclear relaxation is almost invariably studied by magnetic resonance; the weak nuclear paramagnetism cannot be detected by other methods. *See* MAGNETIC RESONANCE.

If the bulk nuclear magnetization **M** is tilted away from the direction of the static field **H**, the **M** vector precesses about **H**. The component of **M** parallel to **H** returns to its thermal equilibrium value by means of spin-lattice relaxation. The spin-spin coupling gives a spread in precession frequencies, causing the nuclei to get out of step with one another, as manifested by the decay of the components of **M** perpendicular to **H** without energy exchange with the lattice (spin-spin relaxation).

Rapid motion of a nucleus relative to its neighbors lengthens the spin-spin relaxation. The times range from tens of microseconds, in almost all solids, to several seconds, in cases of rapid nuclear motion. Measurement of spin-spin relaxation has been extensively used to study atomic motion.

Spin-lattice relaxation arises from magnetic dipole coupling with electrons in the case of metals, paramagnetic substances, or diamagnetic insulators possessing paramagnetic impurities; with other nuclear moments when the relative nuclear positions undergo large changes due to sufficiently rapid self-diffusion or molecular rotation in liquids, gases, or even some solids; or from the nuclear electric quadrupole coupling to the electric fields of the host material. Spin-lattice relaxation times range from less than 1 msec to several hours.

Paramagnetic relaxation. Resonance methods can be used for the study of paramagnetic relaxation. However, the observation of the ability of the magnetization to follow a changing field and even the observation of the change in the lattice temperature brought about by spin-lattice relaxation have also been used. The first interest in magnetic

relaxation, in fact, stemmed from the desire to cool the lattice by adiabatic demagnetization of the spins of paramagnetic ions. Large electrostatic interactions between the ions and their surroundings make major contributions to the total spin energy, and it is rarely possible to neglect this effect. Other important contributions to the spin energy arise from dipolar coupling to applied fields, spin-spin coupling by both magnetic dipole and exchange mechanisms, and magnetic coupling to nuclei, particularly the nuclei of the paramagnetic ions.

The principal spin-lattice relaxation mechanisms involve modulation of the spin-orbit coupling by lattice motion, modulation of the magnetic coupling to nuclei by lattice motion, and coupling with paramagnetic impurity atoms. Typical spin-lattice relaxation times vary from less than 10^{-8} sec at room temperature to tens of seconds at liquid helium temperature. Although the study of paramagnetic relaxation was begun in the 1930s, it is not as thoroughly understood as the more recently studied nuclear relaxation, primarily because of the complexity of the electronic energy levels of the paramagnetic ions.

Ferromagnetic relaxation. In ferromagnetic substances, strong exchange coupling between electron spins, causing them to prefer an alignment parallel to one another, and demagnetizing effects cause ferromagnetic relaxation to differ from paramagnetic relaxation.

The strong tendency of ferromagnetic spins to remain parallel makes it convenient to distinguish between two types of relaxation: (1) that in which the total magnetization changes direction but not magnitude, thereby keeping the exchange energy constant, and (2) that in which the magnitude of the magnetization changes, producing a change in the exchange and also in the demagnetizing energies. The second mechanism depends on the sample shape and may arise from irregularities in the ferromagnetic lattice (impurities and nonstochiometric composition, as in ferrites) as well as from thermal vibrations. Ferromagnetic relaxation has been studied primarily by magnetic resonance.

[C.P.SL.]

Bibliography: E. Abrahams, Relaxation processes in ferromagnetism, *Advances in Electronics and Electron Phys.*, 6:47–68, 1954; A. H. Cooke, Paramagnetic relaxation effects, *Repts. Progr. Phys.*, 13:276–294, 1950; A. J. Dekker, *Solid State Physics*, 1957; C. J. Gorter, *Paramagnetic Relaxation*, 1947.

Magnetic resonance

A phenomenon exhibited by the magnetic spin systems of certain atoms whereby the spin systems absorb energy at specific (resonant) frequencies when subjected to alternating magnetic fields. The magnetic fields must alternate in synchronism with natural frequencies of the magnetic system. In most cases the natural frequency is that of precession of the bulk magnetic moment **M** of constituent atoms or nuclei about some magnetic field **H**. Because the natural frequencies are highly specific as to their origin (nuclear magnetism, electron spin magnetism, and so on) the resonant method enables one to study selectively particular features of interest. For example, one may study weak nuclear magnetism unmasked by the much larger electronic paramagnetism or diamagnetism which usually accompanies it.

Nuclear magnetic resonance (that is, resonance exhibited by nuclei) reveals not only the presence of a nucleus such as hydrogen, which possesses a magnetic moment, but also its interaction with nearby nuclei. It has therefore become a most powerful method of determining molecular structure. The detection of resonance displayed by unpaired electrons, called electron paramagnetic resonance, is also an important application. These two phenomena, as well as other related resonance phenomena, are discussed in this article. It is presumed that the reader has some familiarity with the theory of magnetism (*see* MAGNETISM and the articles listed therein).

Origin of magnetic resonance. Because **M** has its origin in circulating currents or intrinsic spins, there is always an angular momentum **J** associated with it. The vector quantities **M** and **J** are related by the equation

$$\mathbf{M} = \gamma \mathbf{J} \qquad (1)$$

where γ is called the gyromagnetic ratio.

For a system to exhibit a magnetic resonance it must possess a magnetic moment, possess angular momentum, and experience torques. The magnetic moment may arise from nuclei of atoms, from orbital electronic motion, from electronic spins, or from moving nuclear charges during molecular rotation. The angular momentum arises from the same sources. The torques may arise from externally applied magnetic fields; from magnetic dipole fields exerted by neighboring nuclei, atoms, or molecules; from electric fields acting on, for example, a nuclear electric quadrupole moment or the nonspherical electron cloud of an atom; or from electron exchange coupling. In any given case one must decide which interactions are large and which are small. Thus for a paramagnetic atom possessing a nuclear magnetic moment one might distinguish between small applied static fields, in which the coupled nucleus and electron angular momenta act as a unit, and large magnetic fields, which decouple them so that they act independently. An effective angular momentum and magnetic moment can often be defined, giving one an effective γ (in analogy to the Landé *g*-factor of optical spectroscopy).

Several types of resonances have been observed; these differ in one or more of the three basic requirements. However, the principal features can be understood by assuming that the torques arise from an effective static applied magnetic field **H**. The torque $\mathbf{M} \times \mathbf{H}$ causes the angular momentum to change

time according to the equation

$$\frac{d\mathbf{J}}{dt} = \mathbf{M} \times \mathbf{H} \quad \text{or} \quad \frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{H} \qquad (2)$$

The resultant motion of **M** is a precession at angular frequency γH about the direction of **H**. At thermal equilibrium **M** is parallel to **H**, and no precession occurs. Application of an alternating magnetic field $H_x \cos \omega t$ perpendicular to **H** causes **M** to tilt away from **H** with a consequent absorption of energy, provided the resonant condition $\omega = \gamma H$ is satisfied. In practice the absorption takes place over a narrow range of frequency on both sides of γH. The magnetization M_x parallel to H_x obeys the equation

$$M_x = H_x\chi'(\omega) \cos \omega t + H_x\chi''(\omega) \sin \omega t \qquad (3)$$

where $\chi'(\omega)$ and $\chi''(\omega)$ are the real and imaginary parts of the complex magnetic susceptibility $\chi = \chi'(\omega) - j\chi''(\omega)$ $(j = \sqrt{-1})$, and characterize dispersion and absorption respectively. For a typical resonance, $\chi''(\omega)$ attains maximum value for a region of frequencies near $\omega = \gamma H$.

For many cases one must analyze magnetic resonance by quantum theory. Consider a system composed of many (weakly interacting) identical parts (atoms or nuclei) each with angular momentum quantum number F (total angular momentum $= \sqrt{F(F+1)}\hbar$, where $\hbar$ is Planck's constant h divided by 2π). The spatial quantization in the magnetic field H gives $2F + 1$ equally spaced energy levels labeled by the quantum number $M_F = (F, F-1, \ldots, -F)$ and energy spacing between adjacent levels of $\gamma\hbar H$. The field $H_x \cos \omega t$ produces transitions with the selection rule $\Delta M_F = \pm 1$. To satisfy the Bohr frequency condition, $\hbar\omega = \gamma\hbar H$, in agreement with the classical result, $\omega = \gamma H$.

According to quantum theory, the probability of transition from any energy level A to any other B is the same as that from B to A, thus a net absorption of energy requires that the population of the lower energy states be greater than that of the upper (*see* QUANTUM THEORY, NONRELATIVISTIC). The reverse situation in which the upper states are more populated leads to an induced emission and is the basis for the operation of the solid-state maser (*see* MASER). At thermal equilibrium, as a result of spin-lattice relaxation (*see* MAGNETIC RELAXATION) the lower energy states are more heavily populated in accordance with the classical Maxwell-Boltzmann statistics (ordinarily one need not use either Fermi-Dirac or Bose-Einstein statistics). When H_x becomes [illegible] ... [illegible] decreases, a phenomenon known as saturation, because it causes χ' and χ'' to diminish with increasing H_x. Population differences between pairs of states other than A and B may simultaneously be increased, or even inverted, as in the three-level maser. The intensity of the alternating field necessary to produce saturation depends on the width of the absorption line and on the spin-lattice relaxation time (wider lines or shorter times require larger H_x).

Observation of magnetic resonance. Experimentally one can detect magnetic resonance by measuring the absorption of magnetic energy of a circuit containing the magnetic material or by measuring the change in inductance or resonant frequency of the circuit. The two methods measure $\chi''(\omega)$ or $\chi'(\omega)$ respectively. The resonant condition $\omega = \gamma H$ may be produced by varying ω, or more customarily, by changing H. In some experiments one tilts **M** away from the direction of **H** by alternating fields of short duration and then observes voltages induced by the subsequent free precession of **M**. This method is particularly useful for studying relaxation times.

Nuclear magnetic resonance (NMR). The nuclei of many atoms possess angular momentum (spin) and nonvanishing magnetic moments (*see* NUCLEAR MOMENTS). The former may be characterized by an angular momentum quantum number I (integer or half integer) of the nuclear particles. As far as is known, stable nuclei with an even number of neutrons and even number of protons have zero spin and magnetic moment, hence are incapable of exhibiting magnetic resonance.

Nuclear resonances have been observed in insulators, metals, paramagnetic salts, and antiferromagnetic substances and in gases, liquids, and other solids. Often, to observe NMR, a sample is placed between the poles of an electromagnet (Fig. 1) which in addition to the main winding carries a small auxiliary winding or sweep. A coil connected to an oscillator surrounds the sample, as does a second coil at right angles both to the oscillator coil and the sweep winding, to avoid direct coupling. The oscillator frequency is fixed and the sweep circuit is used to vary the magnetic field strength continuously. When a resonance frequency of the sample is reached, a signal induced in the second coil is detected and amplified. Typical resonance frequencies in a field of 10,000 gauss lie

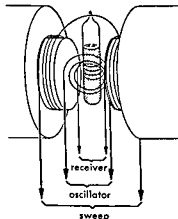


Fig. 1. Arrangement of sample and coils in nuclear induction apparatus. (*From J. D. Roberts, Nuclear Magnetic Resonance, McGraw-Hill, 1959*)

in the radio-frequency region (1–45 megacycles per second, Mc/sec). For example, the H^1 nucleus shows a resonance frequency of 42.6 Mc/sec at this field strength. C^{13} nuclei give a much weaker signal, further decreased in a sample containing this isotope in its natural abundance of 1.1%, and their detection is near to the limit of sensitivity obtained by 1959.

Nuclei with quadrupole moments. If a nucleus has a spin $\geqq 1$ it in general has a nonvanishing electric quadrupole moment (there are good grounds for believing that all nuclei have zero electric dipole moments). The electrical interaction between the nucleus and electric potentials $V(x,y,z)$ from other charges depends on the nuclear orientation (specified by the direction of nuclear spin) and on the spatial second derivatives of the potential $\partial^2V/\partial x^2$, $\partial^2V/\partial x\,\partial y$, and so on, at the position of the nucleus. For potentials of spherical, tetrahedral, or cubic symmetry, the interaction energy is independent of orientation and may be disregarded. When the electric quadrupole interaction is nonzero but nevertheless much weaker than the static magnetic interaction, the unique resonance condition $\omega = \gamma H$ is changed, and the resonance line splits into $2I$ components centered about γH. A convenient method of determining the nuclear spin is thereby provided. The name pure quadrupole resonance is used when one dispenses with the static field H and observes the reorientation of the nucleus among its various quantized orientations with respect to the electric potential alone. Classically, the nonspherical nuclear charge experiences a torque which causes precession.

Applications. Nuclear magnetic resonance has been used widely to measure nuclear magnetic moments, electric quadrupole moments, and spins. Because the resonance lines may be on occasion very sharp (1 cycle wide at 40 Mc), nuclear resonance is frequently employed to measure magnetic fields with great precision. For example, *see* MAGNETOMETER.

The extensive use of NMR in molecular structure determinations arises from the slight shift in the resonance frequency of an atom—commonly that of a proton—due to the environment of neighboring atoms. Because the magnitude of this shift depends on the type of environment, it is called the chemical shift. Figure 2 shows the NMR spectrum of ethyl alcohol, CH_3CH_2OH. The three main resonance frequencies are due to protons in the OH, CH_2, and CH_3 groups, respectively, and the spacing between them (which varies with the field strength) shows the chemical shift characteristic of the protons in these three typical structural groups.

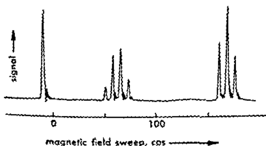


Fig. 2. Proton resonance spectra of ethyl alcohol at 40 Mc/sec. (From J. D. Roberts, Nuclear Magnetic Resonance, McGraw-Hill, 1959)

The separate peaks at each frequency are due to spin-spin splitting. Often, the presence of n protons will split the frequency of a given, structurally different proton into $n + 1$ peaks, in direct analogy to ordinary spectral lines (*see* MOLECULAR STRUCTURE AND SPECTRA). The triplet in the NMR spectrum of ethyl alcohol results from the splitting of the frequency of the CH_3 protons by the two adjacent CH_2 protons; the quadruplet is at the typical CH_2 frequency and is split into four peaks by the three protons of the CH_3 group.

Because the time required for nuclear transitions is relatively long, substances undergoing fast reactions show altered NMR spectra, and rates of such rapid processes as ionization or intramolecular rotation can be measured.

Paramagnetic resonance. Magnetic resonance arising from electrons in paramagnetic substances or from electrons in paramagnetic centers in diamagnetic substances is called *paramagnetic resonance*. For applied fields of several thousand gauss the electron paramagnetic resonance (EPR) experiments are done at microwave frequencies, commonly at 3-cm or at 1-cm wavelengths. In some instances, nuclear resonance apparatus has been used with correspondingly lower applied fields. The most sensitive apparatus at the time this is written detects approximately 10^{12} electron spins for a line 1 gauss broad, a sensitivity far greater than obtained by nonresonant methods, such as those utilizing paramagnetic susceptibility.

Resonances have been observed in atoms of the iron group, rare earths, and other transition elements, in paramagnetic gases, in organic free radicals, in color centers in crystals (such as *F*- and *V*-centers), in metals (conduction electron spin) and in semiconductors (both conduction electron and impurity center spins).

When two paramagnetic ions or molecules approach one another, the spins become coupled via the exchange interaction. Much weaker couplings than the usual exchange interactions within atoms, in chemical bonds, or in ferromagnets produce pronounced effects. For additional information on this phenomenon and its applications, *see* ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY.

Ferromagnetic resonance. In the case of both nuclear and paramagnetic resonance, the spins of neighboring atoms are nearly randomly oriented with respect to one another. In contrast, the electron spins in one domain of a ferromagnet are nearly all parallel for temperatures sufficiently below the Curie point. The alignment may be described in terms of the exchange coupling between neighboring spins, or equivalently in terms of the Weiss molecular magnetic field H_w.

time according to the equation

$$\frac{d\mathbf{J}}{dt} = \mathbf{M} \times \mathbf{H} \quad \text{or} \quad \frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{H} \qquad (2)$$

[illegible] precession at angular [illegible] occurs. Application of an alternating magnetic field $H_x \cos \omega t$ perpendicular to **H** causes **M** to tilt away from **H** with a consequent absorption of energy, provided the resonant condition $\omega = \gamma H$ is satisfied. In practice the absorption takes place over a narrow range of frequency on both sides of γH The magnetization M_x parallel to H_x obeys the equation

$$M_x = H_x\chi'(\omega) \cos \omega t + H_x\chi''(\omega) \sin \omega t \qquad (3)$$

where $\chi'(\omega)$ and $\chi''(\omega)$ are the real and imaginary parts of the complex magnetic susceptibility $\chi = \chi'(\omega) - j\chi''(\omega)$ $(j = \sqrt{-1})$, and characterize dispersion and absorption respectively. For a typical resonance, $\chi''(\omega)$ attains maximum value for a region of frequencies near $\omega = \gamma H$.

For many cases one must analyze magnetic resonance by quantum theory Consider a system composed of many (weakly interacting) identical parts (atoms or nuclei) each with angular momentum quantum number F (total angular momentum = $\sqrt{F(F+1)}\hbar$, where $\hbar$ is Planck's constant h divided by 2π). The spatial quantization in the magnetic field H gives $2F + 1$ equally spaced energy levels labeled by the quantum number $M_F = (F, F-1, \ldots, -F)$ and energy spacing between adjacent levels of $\gamma\hbar H$ The field $H_x \cos \omega t$ produces transitions with the selection rule $\Delta M_F = \pm 1$. To satisfy the Bohr frequency condition, $\hbar\omega = \gamma\hbar H$, in agreement with the classical result, $\omega = \gamma H$.

According to quantum theory, the probability of transition from any energy level A to any other B is the same as that from B to A; thus a net absorption of energy requires that the population of the lower energy states be greater than that of the upper (*see* QUANTUM THEORY, NONRELATIVISTIC). The reverse situation in which the upper states are more populated leads to an induced emission and is the basis for the operation of the solid-state maser (*see* MASER). At thermal equilibrium, as a result of spin-lattice relaxation (*see* MAGNETIC RELAXATION) the lower energy states are more heavily populated in accordance with the classical Maxwell-Boltzmann statistics (ordinarily one need not use either Fermi-Dirac or Bose-Einstein statistics). When H_x becomes [illegible] decreases, a phenomenon known as saturation, because it causes χ' and χ'' to diminish with increasing H_x. Population differences between pairs of states other than A and B may simultaneously be increased, or even inverted, as in the three-level maser. The intensity of the alternating field necessary to produce saturation depends on the width of the absorption line and on the spin-lattice relaxation time (wider lines or shorter times require larger H_x).

Observation of magnetic resonance. Experimentally one can detect magnetic resonance by measuring the absorption of magnetic energy of a circuit containing the magnetic material or by measuring the change in inductance or resonant frequency of the circuit. The two methods measure $\chi''(\omega)$ or $\chi'(\omega)$ respectively. The resonant condition $\omega = \gamma H$ may be produced by varying ω, or more customarily, by changing H. In some experiments one tilts **M** away from the direction of **H** by alternating fields of short duration and then observes voltages induced by the subsequent free precession of **M**. This method is particularly useful for studying relaxation times.

Nuclear magnetic resonance (NMR). The nuclei of many atoms possess angular momentum (spin) and nonvanishing magnetic moments (*see* NUCLEAR MOMENTS). The former may be characterized by an angular momentum quantum number I (integer or half integer) of the nuclear particles. As far as is known, stable nuclei with an even number of neutrons and even number of protons have zero spin and magnetic moment, hence are incapable of exhibiting magnetic resonance.

Nuclear resonances have been observed in insulators, metals, paramagnetic salts, and antiferromagnetic substances and in gases, liquids, and other solids. Often, to observe NMR, a sample is placed between the poles of an electromagnet (Fig. 1) which in addition to the main winding carries a small auxiliary winding or sweep. A coil connected to an oscillator surrounds the sample, as does a second coil at right angles both to the oscillator coil and the sweep winding, to avoid direct coupling. The oscillator frequency is fixed and the sweep circuit is used to vary the magnetic field strength continuously. When a resonance frequency of the sample is reached, a signal induced in the second coil is detected and amplified. Typical resonance frequencies in a field of 10,000 gauss lie

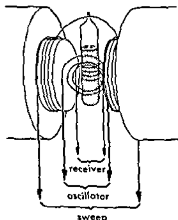


Fig. 1. Arrangement of sample and coils in nuclear induction apparatus. (*From J. D. Roberts, Nuclear Magnetic Resonance, McGraw-Hill, 1959*)

in the radio-frequency region (1–45 megacycles per second, Mc/sec). For example, the H^1 nucleus shows a resonance frequency of 42.6 Mc/sec at this field strength. C^{13} nuclei give a much weaker signal, further decreased in a sample containing this isotope in its natural abundance of 1.1%, and their detection is near to the limit of sensitivity obtained by 1959.

Nuclei with quadrupole moments. If a nucleus has a spin $\geqq 1$ it in general has a nonvanishing electric quadrupole moment (there are good grounds for believing that all nuclei have zero electric dipole moments). The electrical interaction between the nucleus and electric potentials $V(x,y,z)$ from other charges depends on the nuclear orientation (specified by the direction of nuclear spin) and on the spatial second derivatives of the potential $\partial^2 V/\partial x^2$, $\partial^2 V/\partial x\,\partial y$, and so on, at the position of the nucleus. For potentials of spherical, tetrahedral, or cubic symmetry, the interaction energy is independent of orientation and may be disregarded. When the electric quadrupole interaction is nonzero but nevertheless much weaker than the static magnetic interaction, the unique resonance condition $\omega = \gamma H$ is changed, and the resonance line splits into $2I$ components centered about γH. A convenient method of determining the nuclear spin is thereby provided. The name *pure quadrupole resonance* is used when one dispenses with the static field H and observes the reorientation of the nucleus among its various quantized orientations with respect to the electric potential alone. Classically, the nonspherical nuclear charge experiences a torque which causes precession.

Applications. Nuclear magnetic resonance has been used widely to measure nuclear magnetic moments, electric quadrupole moments, and spins. Because the resonance lines may be on occasion very sharp (1 cycle wide at 40 Mc), nuclear resonance is frequently employed to measure magnetic fields with great precision. For example, *see* MAGNETOMETER.

The extensive use of NMR in molecular structure determinations arises from the slight shift in the resonance frequency of an atom—commonly that of a proton—due to the environment of neighboring atoms. Because the magnitude of this shift depends on the type of environment, it is called the chemical shift. Figure 2 shows the NMR spectrum of ethyl alcohol, CH_3CH_2OH. The three main resonance frequencies are due to protons in the OH, CH_2, and CH_3 groups, respectively, and the spacing between them (which varies with the field strength) shows the chemical shift characteristic of the protons in these three typical structural groups.

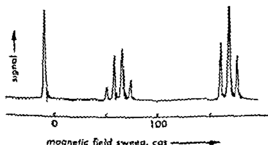


Fig. 2. Proton resonance spectra of ethyl alcohol at 40 Mc/sec. (From J. D. Roberts, *Nuclear Magnetic Resonance*, McGraw-Hill, 1959)

The separate peaks at each frequency are due to spin-spin splitting. Often, the presence of n protons will split the frequency of a given, structurally different proton into $n + 1$ peaks, in direct analogy to ordinary spectral lines (*see* MOLECULAR STRUCTURE AND SPECTRA). The triplet in the NMR spectrum of ethyl alcohol results from the splitting of the frequency of the CH_3 protons by the two adjacent CH_2 protons; the quadruplet is at the typical CH_2 frequency and is split into four peaks by the three protons of the CH_3 group.

Because the time required for nuclear transitions is relatively long, substances undergoing fast reactions show altered NMR spectra, and rates of such rapid processes as ionization or intramolecular rotation can be measured.

Paramagnetic resonance. Magnetic resonance arising from electrons in paramagnetic substances or from electrons in paramagnetic centers in diamagnetic substances is called paramagnetic resonance. For applied fields of several thousand gauss the electron paramagnetic resonance (EPR) experiments are done at microwave frequencies, commonly at 3-cm or at 1-cm wavelengths. In some instances, nuclear resonance apparatus has been used with correspondingly lower applied fields. The most sensitive apparatus at the time this is written detects approximately 10^{12} electron spins for a line 1 gauss broad, a sensitivity far greater than obtained by nonresonant methods, such as those utilizing paramagnetic susceptibility.

Resonances have been observed in atoms of the iron group, rare earths, and other transition elements, in paramagnetic gases, in organic free radicals, in color centers in crystals (such as *F*- and *V*-centers), in metals (conduction electron spin) and in semiconductors (both conduction electron and impurity center spins).

When two paramagnetic ions or molecules approach one another, the spins become coupled via the exchange interaction. Much weaker couplings than the usual exchange interactions within atoms, in chemical bonds, or in ferromagnets produce pronounced effects. For additional information on this phenomenon and its applications, *see* ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY.

Ferromagnetic resonance. In the case of both nuclear and paramagnetic resonance, the spins of neighboring atoms are nearly randomly oriented with respect to one another. In contrast, the electron spins in one domain of a ferromagnet are nearly all parallel for temperatures sufficiently below the Curie point. The alignment may be described in terms of the exchange coupling between neighboring spins, or equivalently in terms of the Weiss molecular magnetic field H_w.

It is simplest to consider the case where magnetization is uniform throughout the sample. Neglecting relaxation effects, the equation of motion is still Eq. (2), but an effective field is substituted for **H**, consisting of the applied static and alternating fields, the demagnetizing corrections (from the electron magnetic dipolar fields), and the effects of crystalline anisotropy. Because the Weiss molecular field is always parallel to **M**, $\mathbf{H}_w$ exerts no torque and plays no role as long as the magnetization is uniform throughout the sample. (The exchange energy between spins does not change as long as their relative orientation does not change.)

The crystalline anisotropy can be shown to be equivalent to a magnetic field $\mathbf{H}_A$ along the direction of easy magnetization as long as **M** points nearly in that direction.

Because of the demagnetizing effects, the resonant frequency depends on sample geometry. For an infinite plane perpendicular to the applied field, the resonant angular frequency ω is given by $\omega = \gamma\sqrt{BH}$, but for a sphere, $\omega = \gamma H$.

The large demagnetizing and exchange fields are the principal difference between ferromagnetic and paramagnetic resonance. The demagnetizing field has components $-N_xM_x$, $-N_yM_y$, and $-N_zM_z$ [illegible] ing field tilts **M** away from z, changing the effective z field. If the change brings the spins closer to resonance, **M** may tilt out more. It is possible for such a nonlinear effect to be unstable for sufficiently large alternating fields. This instability is utilized in the Suhl ferromagnetic amplifier.

Antiferromagnetic resonance. The two sublattices of spins in an antiferromagnet are strongly coupled together by exchange forces. If one tilts both magnetizations ($\mathbf{M}_1$ and $\mathbf{M}_2$) together, away from the normal direction of magnetization in the crystal (call this the z direction), the only change in energy results from the anisotropy fields. However, because the anisotropy fields are reversed in direction for the two lattices, the magnetizations $\mathbf{M}_1$ and $\mathbf{M}_2$ tend to precess in opposite directions, bringing about a change in the exchange energy. An external field along the z direction aids one anisotropy field but opposes the other. The resonant angular frequency ω for a sphere is given by

$$\omega = \gamma(H \pm \sqrt{H_A(H_A + 2H_E)}) \qquad (4)$$

where H is the applied field, H_A the equivalent anisotropy field, and H_E the equivalent exchange field. The plus and minus signs refer to two opposite directions of rotating magnetic fields which may be used to observe the resonance. If H_E is 10^6 oersted, and H_A is 10^4 oersted, the corresponding frequency is 3×10^{11} cps.

Ferrimagnetic resonance. Magnetic resonance in ferrites is called ferrimagnetic resonance. Ferrites are the natural generalization of antiferromagnets, containing two or more sublattices which may differ in magnetization. The basic coupling terms are still anisotropy fields, exchange fields, and the applied fields. The resonant angular frequency ω for the case of two sublattices is

$$\omega = \gamma\left[H - \frac{\eta H_E}{2} \pm \sqrt{\left(\frac{\eta H_E}{2}\right)^2 + H_EH_A(2-\eta) + H_A^2}\right] \qquad (5)$$

where η is a parameter measuring the relative sizes of the two magnetization vectors. Taking $\mathbf{M}_1$ to be the smaller magnetization, η is defined by

$$\mathbf{M}_1 = (1-\eta)\mathbf{M}_2 \qquad (6)$$

Equation (6) assumes the magnetizations to be at saturation and the two sublattices to have the same γ (deviations might differ from spin-orbit coupling). See MOLECULAR BEAMS. [C.P.SL.]

Bibliography: A. J. Dekker, *Solid State Physics*, 1957; G. E. Pake, *Magnetic resonance*, *Sci. Ameri-* [illegible] Schneider, and H. J. Bernstein, *High Resolution Nuclear Magnetic Resonance*, 1960; J. D. Roberts, *Nuclear Magnetic Resonance*, 1959; F. Seitz and D. Turnbull (eds.), *Solid State Physics*, vol. 2, 1956.

Magnetic separation methods

All materials possess magnetic properties. Substances that have a greater permeability than air are classified as paramagnetic; those with a lower permeability are called diamagnetic. Paramagnetic materials are attracted to a magnet; diamagnetic substances are repelled. Very strongly paramagnetic materials are classified as ferromagnetic and include such metals as iron, nickel, and cobalt, and such minerals as magnetite, pyrrhotite, and ilmenite. Such substances can be separated from weakly or nonmagnetic materials by the use of low-intensity magnetic separators. Minerals such as hematite, limonite, and garnet are weakly magnetic and can be separated from nonmagnetics by the use of high-intensity separators.

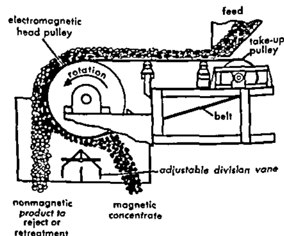


Fig. 1. Belt type magnetic cobber. (*Stearns Magnetic Products*)

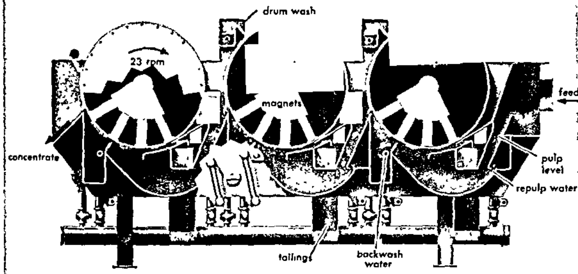


Fig. 2. Countercurrent wet magnetic separator. (*Jeffrey Manufacturing Co.*)

Magnetic separators are widely used to remove tramp iron from ores being crushed so as to protect crushers, to remove contaminating magnetics from food and industrial products, to recover magnetite and ferrosilicon in the float-sink methods of ore concentration, and to upgrade or concentrate ores. Tramp iron may be removed by a stationary magnet suspended over a conveyor belt or the material may be passed over a magnetic pulley; in both cases, magnetics are attracted by the magnetic field and separated (Fig. 1).

Magnetic separators are extensively used to concentrate ores, particularly iron ores, when one of the principal constituents is magnetic. When the chief economic mineral is magnetite, iron ores can be cheaply and effectively separated by low intensity separators. Such separators may be dry or wet. If an ore can be crushed to give substantial liberation of minerals at sizes coarser than ¼ in., separations can be made on the +¼-in. sizes on a belt-type magnetic cobber (Fig. 1). Such a machine is usually used to cob out or reject waste materials. If clean mineral occurs at this size it can be recovered by splitting the discharge or by retreatment on a lower strength machine. Wet magnetic separators are usually used to treat ore finer than ⅛ in. These may be of the belt type or of the more common rotating drum type. Drum-type separators consist of one or more rotatable drums having inner nonrotatable magnet elements with 3–7 poles. The magnets may be either electromagnets or permanent magnets. The feed enters the machine as a slurry; the magnetics are attracted to the pole pieces and are carried to a discharge point on the surface of the drum. Many types of box designs are in use. The concurrent type is frequently used on relatively coarse material to reject a clean waste product. The countercurrent type (Fig. 2) is employed on fine ore to give a clean concentrate. Magnets may be of either the electromagnetic or permanent type. Electromagnets were formerly used almost exclusively but are now used mainly when exceptionally high field strengths are required or when it is desirable to vary magnet strength. Permanent magnets are now widely used since modern materials permit charging and retaining high field strengths permanently. Most permanent magnets are of the Alnico type but the ceramic types containing barium ferrites are coming into use. Several types of magnetic separators employing alternating current have been devised but have found little commercial use.

High-intensity separators for the separation of weakly magnetic minerals are usually of the dry type. Surface-tension effects usually rule out wet separations. Since magnetic attraction varies inversely as the square of the distance, weakly magnetic minerals must be brought close to the magnets if they are to be separated. Both belt-type and induced-roll machines are used. The ore must be thoroughly dried and closely sized for best performance. *See* MAGNETIC MATERIALS; MAGNETISM; ORE DRESSING; SEPARATION (CHEMICAL and PHYSICAL); SEPARATION (MECHANICAL). [F.D.D.]

Bibliography: R. S. Dean and C. W. Davis, *Magnetic Separation of Ores*, U.S. Bur. Mines Bull. 425, 1941; L. A. Roe, *Iron Ore Beneficiation*, 1957.

Magnetism

Magnetism comprises those physical phenomena involving magnetic fields and their effects upon materials. Magnetic fields may be set up on a macroscopic scale by electric currents or by magnets. On an atomic scale, individual atoms cause magnetic fields when their electrons have a net magnetic moment as a result of their angular momentum. A magnetic moment arises whenever a charged particle has an angular momentum. It is the cooperative effect of the atomic magnetic moments which causes the macroscopic magnetic field of a permanent magnet. *See* ELECTRON SPIN; MAGNETIC MOMENT.

History. The history of magnetism reaches back into ancient times. The name magnet was used by the Greeks for a stone which was capable of attracting other pieces of the same material and iron as well. This original magnet (lodestone) is the naturally occurring magnetic iron oxide, magnetite. Some of the properties of magnetite were discovered earlier than 600 B.C., although it is only in the twentieth century that physicists have begun to understand why substances behave magnetically. Magnetism is one of the earliest known physical phenomena of solid materials.

William Gilbert (1540–1603) was the first to apply scientific methods to a systematic exploration of magnetic phenomena. His greatest contribution was the discovery that the earth is itself a magnet (*see* GEOMAGNETISM). He distinguished clearly between electricity and magnetism. Quantitative studies of magnetism began in the early eighteenth century. Charles Coulomb (1736–1806) established the inverse square law of force between magnetic poles and also between electric charges (*see* COULOMB'S LAW). S. D. Poisson (1781–1840) set up the basis of magnetostatics through application of potential theory to the problem of the forces between magnetized bodies. In the nineteenth century, James Clerk Maxwell (1831–1879) established the relationship between magnetism and electricity by developing the science of electromagnetism (*see* ELECTRICITY; ELECTROMAGNETISM; MAGNETOSTATICS).

Magnetic field. A magnetic field is said to occupy a region when the magnetic effect of an electric current or of a magnet upon a small test magnet which is brought in the vicinity is detectable. Because of the magnetic effect, a torque will be exerted on the test magnet until it becomes oriented in a particular direction. The magnitude of this torque is a measure of the strength of the magnetic field, and the preferred direction of orientation is the direction of the field. In electromagnetic units, the magnitude of the torque is given by

$$L = -|\mathbf{\mu}||\mathbf{H}| \sin \theta$$

where $\mathbf{\mu}$ is the magnetic moment of the test magnet, $\mathbf{H}$ is the magnetic field strength, and θ is the angle between the direction of $\mathbf{\mu}$ and the direction of $\mathbf{H}$.

When a magnetic material is placed in a magnetic field $\mathbf{H}$, it becomes magnetized; that is, it becomes itself a magnet. The intensity of this induced magnetism is called the magnetization $\mathbf{M}$. More precisely, $\mathbf{M}$ is the magnetic moment per unit volume of the material. A vector field, the magnetic induction $\mathbf{B}$, is often defined to describe the magnetic forces anywhere in space. In electromagnetic units, the definition of $\mathbf{B}$ is

$$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$$

See INDUCTION, MAGNETIC; MAGNETIC FIELD; MAGNETIC FLUX; MAGNETIZATION.

Classification of substances. Materials can be grouped as to their behavior in magnetic fields:

1. Diamagnetic substances have a negative magnetic susceptibility; they are magnetized in a direction opposite to that of an applied magnetic field. *See* DIAMAGNETISM; SUSCEPTIBILITY, MAGNETIC.

2. Paramagnetic substances have a positive magnetic susceptibility such that their magnetization is parallel and usually proportional to the applied magnetic field. *See* PARAMAGNETISM.

3. A ferromagnetic substance is one in which, below a certain temperature, the atomic magnetic moments tend to line up and point in a common direction. Above this temperature (the Curie temperature), the substance behaves as if it were paramagnetic. The cooperative effect of the atomic magnetic moments gives rise to a relatively large magnetization for a given applied field. *See* CURIE TEMPERATURE, MAGNETIC; FERRIMAGNETISM; FERROMAGNETISM.

4. An antiferromagnetic substance is one in which, below a certain temperature, the atomic magnetic moments point alternatingly in opposite directions in an ordered manner. *See* ANTIFERROMAGNETISM; *see also* LOW-TEMPERATURE PHYSICS; MAGNETIC MATERIALS; MAGNETIC RESONANCE; PERMEABILITY, MAGNETIC. [E.A.; F.K.]

Bibliography: L. F. Bates, *Modern Magnetism*, 3d ed., 1951; D. Shoenberg, *Magnetism*, 1949; E. C. Stoner, *Magnetism and Matter*, 1934.

Magnetite

An iron-ore mineral having composition Fe_3O_4 and crystallizing in the isometric system. Octahedral crystals of magnetite are common but it usually occurs in granular massive aggregates. Octahedral parting is shown by some specimens. The hardness is 6 (Mohs scale) and the specific gravity is 5.18. The luster is metallic and the color iron black. It is strongly magnetic and the variety known as lodestone acts as a natural magnet.

Magnetite is a widespread mineral second only to hematite as an iron ore. It is found as an accessory mineral in most igneous rocks, and in some basic rocks, frequently associated with ilmenite, it has been concentrated in large masses by magmatic segregation. It also occurs in large lenses and beds in metamorphic rocks. An intimate mixture of magnetite and corundum make up most of the material known as emery. Important ore bodies of magnetite are found in Norway, Rumania, and Russia, but the world's largest are in northern Sweden at Kiruna and Gellivare. In the United States, magnetite is mined in the Adirondack region of New York, in Pennsylvania, and in Utah. *See* EMERY; HEMATITE; ILMENITE; IRON (EXTRACTION FROM ORE); ORE AND MINERAL DEPOSITS.

[C.S.HU.]

Magnetization

The process of becoming magnetized; also the property and in particular the extent of being magnetized. Magnetization has an effect on many of the physical properties of a substance. Among these are electrical resistance, specific heat, and elastic strain. *See* MAGNETOCALORIC EFFECT; MAGNETORESISTANCE; MAGNETOSTRICTION.

The magnetization M *of a body is caused by circulating electric currents or by elementary atomic magnetic moments, and is defined as the magnetic moment per unit volume of such currents or moments.* In the electromagnetic system of units (emu), M is measured in gauss; in the mks system, M is measured in webers per square meter (webers/m²). *For M, 1 weber/m² = $10^4/4\pi$ gauss.*

The magnetic induction or magnetic flux density B is given by

$$B = H + 4\pi M \quad \text{(emu)} \qquad (1)$$

where B and M are in gauss, and H, the applied magnetic field, is in oersteds (equivalent to gauss); or by

$$B = \mu_0 H + M \quad \text{(mks)} \qquad (2)$$

where B and M are in weber/m², H is in amp-turns/m, and μ_0, *the permeability of free space*, is defined as $4\pi \times 10^{-7}$ henries/m, that is, webers/(amp-turn)(m). *See* ELECTRICAL UNITS; INDUCTION, MAGNETIC.

The permeability μ of a substance is defined as the ratio B/H (in emu) or $B/\mu_0 H$ (in mks). The magnetic susceptibility χ is defined as the ratio M/H (in emu) or $M/\mu_0 H$ (in mks). From Eqs. (1) and (2):

$$\mu = 1 + (4\pi)\chi \qquad (3)$$

where the 4π is to be used only in the emu system. It is to be noted that the magnitude of μ is the same *in both systems of units, but the magnitude of* χ differs by a factor of 4π. Both μ and χ may be tensors, although usually they are simple numbers. *See* PERMEABILITY, MAGNETIC; SUSCEPTIBILITY, MAGNETIC.

Spontaneous magnetization, or magnetization in the absence of H, is exhibited only by ferromagnetic materials below the Curie temperature; and magnetization is generally limited to ferromagnets. (Antiferromagnets exhibit spontaneous sublattice magnetization; *see* ANTIFERROMAGNETISM.) A ferromagnet is composed of an assemblage of spontaneously magnetized regions called domains. *Within each domain, the elementary atomic magnetic* moments are essentially aligned, that is, each domain may be envisioned as a small magnet. An unmagnetized ferromagnet is composed of numerous domains, oriented in some fashion as shown in Fig. 1, so that the total magnetization is zero.

The process of magnetization in an applied field H consists of growth of those domains oriented

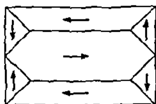


Fig. 1. Possible arrangement of domains in an unmagnetized single crystal. The arrangement is much more haphazard in polycrystals.

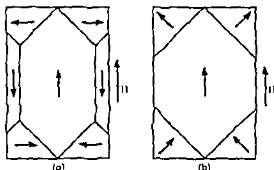


Fig. 2. The process of magnetization. (a) Domain growth. (b) Rotation.

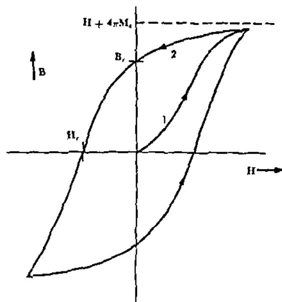


Fig. 3. Magnetization curve.

most nearly in the direction of H at the expense of others (Fig. 2*a*), followed by rotation of the direction of magnetization against anisotropy forces (Fig. 2*b*). For discussions of anisotropy and of the *nature and origin of domains, see* FERROMAGNETISM.

On removal of the field H, some of the magnetization will remain; this is called the remanence M_r.

Magnetization curves. These curves, sometimes called *B-H* curves, are used to describe magnetic materials. They are plotted with H as abscissa and with either M or B as ordinate. In Fig. 3, B_r is the remanent induction ($B_r \approx 4\pi M_r$); H_c is the coercive force, or reverse field required to bring the induction B back to zero; and M_s is the saturation magnetization, or magnetization when all domains are aligned The saturation magnetization is equal to the spontaneous magnetization of a single domain, except that it is possible to increase this magnetization slightly by application of an extremely large field. Saturation magnetization is temperature dependent, and disappears completely above the Curie temperature T_c where a ferromagnet changes into a paramagnet (*see* CURIE TEMPERATURE, MAG-

Values of saturation magnetization M_s and ferromagnetic Curie temperature T_c

Substance	M_s at 293°K, gauss	M_s at 0°K, gauss	T_c, °K
Fe	1707	1752	1013
Co	1400	1446	1388
Ni	485	510	631
Gd	0	1980	289
MnBi	600	675	670
Fe_3O_4	485	?	818

NETIC) The table lists M_s and T_c for a few ferromagnetic materials.

The initial permeability μ_i of a substance is the slope of the magnetization curve at H = 0. There is a definite correlation between initial permeability and coercive force; that is, materials of large μ_i have small H_c and vice versa.

The relationship between B and H may be studied by means of a Rowland ring or toroid in which the core is the material to be studied For this arrangement, the numerical value of H can be computed from the relation

$$H = \frac{NI}{l}$$

where N is the number of turns of the coil, I is the current in the coil, and l is the length of the mean circumference of the coil. If the radius of the ring is large compared to the radial distance across the core, the flux density will be nearly uniform.

The flux density B is not a linear function of H, and furthermore, its value depends not only upon H but upon the previous magnetic history of the sample used.

The ferromagnetic core of the Rowland ring may be demagnetized by successive reversals of the current as the current is gradually reduced from a sufficiently high initial current to a current that will produce a magnetizing force less than any value to be used subsequently. *See* DEMAGNETIZATION.

Changes in B, but not B itself, can be measured by wrapping a search coil on the toroid and observing the throw of a ballistic galvanometer. *See* SEARCH COIL.

Normal magnetization curve. When the core of the Rowland ring is initially demagnetized, B = 0 and H = 0. If the current is quickly increased to give a predetermined value H_1, a flux Φ_1 will be set up in the core, and the change of flux $\Phi_1 - 0$ will be measured by the throw of the ballistic galvanometer. From the flux and the area of A of cross section of the core, the value of $B = \Phi/A$ corresponding to the selected H is computed. Then, without reducing the current, a second increase in current is made to reach H_2. Again, the throw of the galvanometer measures the change in flux $\Phi_2 - \Phi_1$, and hence Φ_2 is found and B_2 is computed corresponding to H_2. By continuing this process for as many pairs of B and H as are necessary, points on the magnetization curve are obtained. The curve obtained starting with the core demagnetized is called the normal magnetization curve (Fig. 4). If the specimen were not initially demagnetized, the process here described would yield a magnetization curve that would differ in shape from the normal magnetization curve and, in general, would not pass through the origin.

In the curve of Fig. 4, B increases slowly at first as H rises, then increases rapidly until the "knee" of the curve is reached. Here, the rate of change of B with respect to H decreases and becomes small as saturation is approached.

Single crystals. Magnetization curves of single crystals depend upon the direction of H with respect to the crystallographic axes. If a weak field is applied to a crystal, those domains whose magnetic moments are most nearly parallel to H increase in size at the expense of those in other directions. There is then a small net contribution of the domains to the flux. As the field increases, the boundaries of the domains continue to change, and B increases rapidly along the steep part of the curve. Near the end of this change, all the magnetic moments in the domains rotate in a direction parallel to those that have been increasing. The rotation within the domains requires greater change in H, and thus the slope changes at the knee. As H is further increased, the domains rotate until their magnetic moments are all parallel to the field and saturation has been reached.

The ferromagnetic materials in common use are polycrystalline, that is, a piece of the material consists of a tremendous number of single crystals of random orientation In each single crystal, there are many domains. The magnetization of the whole body consists of the magnetization of the various single crystals.

Domain growth. The process of magnetization is shown in Fig. 5*a*. Domain growth proceeds by movement of the so-called Bloch wall between domains. This takes place reversibly at first, then irreversibly. Irreversible motion causes sudden, almost discontinuous changes in magnetization, called the Barkhausen effect (Fig. 5*b*). *See* BARKHAUSEN EFFECT.

Hysteresis. The irreversible nature of magnetization is shown most strikingly by the fact that the

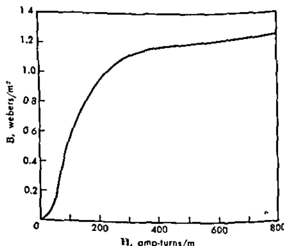


Fig. 4. Normal magnetization curve.

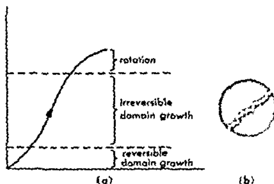


Fig 5. (a) Portion of a magnetization curve, showing magnetization processes. (b) Enlarged section of path, showing Barkhausen effect.

path of demagnetization does not retrace the path of magnetization—the path 2 of Fig. 3 does not retrace path 1. There is a tendency for the magnetization to show hysteresis, that is, to lag behind the applied field, and the loop of Fig. 3 is called a hysteresis loop (*see* HYSTERESIS, MAGNETIC). The area of the loop is a measure of the energy lost to heat, per cycle, in going around the loop, and therefore is related to the energy lost by a magnet in an alternating field. An additional energy loss is that from eddy currents; this is absent in the nonconducting ferrites and may be minimized in metals by lamination.

Demagnetizing fields. The effect of an applied field H on a sample is always reduced by a demagnetizing field $-NM$ coming from the surface "poles," where N, the demagnetizing factor, depends upon the sample shape and orientation with respect to H. It varies from nearly 0 in a sample long and thin in the direction of H, to nearly 4π (in emu) in a sample short and fat, like a disk of revolution about H. Magnetization curves are generally obtained from long, thin samples to avoid demagnetization. The demagnetization effects are extremely important in ferromagnetic resonance (*see* MAGNETIC RESONANCE). [E.A.; F.K.; K.V.M.]

Bibliography: K. Hoselitz, *Ferromagnetic Properties of Metals and Alloys*, 1952; L. Page and N. I. Adams, *Principles of Electricity*, 3d ed., 1958; E. R. Peck, *Electricity and Magnetism*, 1953; R. P. Winch, *Electricity and Magnetism*, 1955.

Magneto

A type of permanent-magnet alternating-current generator frequently used as a source of ignition energy on tractor, marine, industrial, and aviation engines. The higher cost of magneto ignition is not warranted in modern automobiles, where storage batteries are required for other electrically operated equipment. Hand-operated magneto generators were once widely used for signaling from local battery telephone sets. *See* GENERATOR, ELECTRIC.

Multipolar aviation magneto in dual-ignition circuit for nine-cylinder radial engine. (Bendix Aviation Corp)

Modern induction-type magnetos consist of a permanent-magnet rotor and stationary low- and high-tension windings, also called the primary and secondary windings. The illustration shows two induction-type magnetos in a dual-ignition circuit for a nine-cylinder aircraft engine. The two magnetos are identical, but only Magneto no. 1 is shown schematically. The ignition system may receive energy from either or both magnetos, according to the position of the switch.

The energy output of a magneto is obtained as a result of a rapid rate of change of flux through the stationary windings. The primary winding has comparatively few turns and the secondary winding has many thousand turns of fine wire. One end of the secondary winding is connected to an end of the primary winding and grounded to the frame of the magneto The primary winding is closed on itself through a breaker mechanism actuated by a cam on the magneto shaft. The breaker is mechanically set to interrupt the primary circuit each time the flux through the winding is changing at its greatest rate. The sudden collapse of the primary current induces a very high voltage in the secondary winding.

Magnetos are always geared to the engine shaft and timed to open the breaker at the proper instant. Magneto speed depends on the number of poles of the magneto and the number of engine cylinders. The distributor finger speed is always one-half engine speed. *See* IGNITION SYSTEM; INTERNAL COMBUSTION ENGINE. [R.T.W.]

Bibliography: L. S. Marks (ed.), *Mechanical Engineers' Handbook*, 5th ed., 1951; H. Pender and W. A. Del Mar (eds.), *Electrical Engineers' Handbook, Electric Power*, 4th ed., 1949.

Magnetocaloric effect

The reversible change of temperature accompanying the change of magnetization of a ferromagnetic or paramagnetic material. Thermodynamics theory shows that, for an adiabatic change of field ΔH, the change of temperature ΔT is given by

$$\frac{\Delta T}{\Delta H} = -\frac{T}{c_H}\left(\frac{\partial M}{\partial T}\right)_H$$

where c_H is the specific heat per unit volume at constant H, and M is the magnetization.

This change in temperature may be of the order of 1°C, and is not to be confused with the much smaller *hysteresis* heating effect, which is irreversible. *See* HYSTERESIS, THERMAL.

Since $(\partial M/\partial T)_H$ is negative, an adiabatic decrease in H causes T to drop. This is the basis of the Giauque-Debye *adiabatic* demagnetization of paramagnetic salts, a technique which has achieved extremely low *temperatures*. *See* ADIABATIC DEMAGNETIZATION. [E.A.; F.K.]

Magnetochemistry

The branch of chemistry which studies the interrelationship between a *magnetic field* and atomic and molecular structures.

When a substance is placed in a magnetic field of strength H, the magnetic induction is given by B, where

$$B = H + 4\pi I$$

The quantity I is the intensity of magnetization, and $I/H = \kappa$ is the magnetic susceptibility per unit volume. The magnetic susceptibility per unit mass is $\kappa/D = \chi$, where D is the density.

A substance in a magnetic field acquires an intensity of magnetization which may be either smaller or larger than that induced in a vacuum by the same field. In the first case, the substance is said to be diamagnetic. In the second case, the substance may be paramagnetic, ferromagnetic or antiferromagnetic.

Diamagnetism, a universal property of matter, is usually of the order of magnitude 10^{-6}–10^{-5}. Temperature-dependent paramagnetism, on the other hand, arises only when an atom, ion, or molecule possesses a permanent magnetic moment either in the ground state or in an excited state. A permanent magnetic moment is the result of the presence of one or more unpaired electrons. Paramagnetic susceptibilities are of the order of magnitude 10^{-4}–10^{-3}.

A substance composed of atoms with permanent magnetic moments which are very near to one another (for example, iron metal) may display ferromagnetism. This phenomenon occurs when a large number of the atoms with permanent magnetic moments interact so that their individual moments align in a parallel fashion, giving rise to a large resultant moment.

On the other hand, a similar substance (for example, manganese metal) may display antiferromagnetism. Here, the magnetic moments align in an antiparallel fashion, thus largely canceling the individual magnetic moments of the atoms. Parallel versus antiparallel alignment depends, among other factors, upon interatomic distances. Magnetic theories are not yet refined enough to enable one to predict whether a given substance will be ferromagnetic or antiferromagnetic.

In general, the susceptibility of diamagnetic substances is independent of temperature and of field strength. The susceptibility of paramagnetic substances is often inversely proportional to the absolute temperature but independent of field strength. The susceptibility of ferromagnetic and antiferromagnetic substances is dependent on both temperature and field strength in a rather complicated way.

There are many methods available for the measurement of magnetic susceptibility. Most of these methods involve measuring the force exerted on a sample by a magnetic field.

Atomic diamagnetism. The only important application of diamagnetic ionic susceptibilities is their use as correction factors for measured susceptibilities. All substances, even though paramagnetic, have an underlying diamagnetism. A precise determination of the paramagnetic susceptibility

of sodium neptunyl acetate, for example, should include subtraction from the measured molar susceptibility of the diamagnetic ionic susceptibilities of sodium, acetate, and neptunyl ions.

The diamagnetism of atoms and ions can be calculated theoretically by considering electron density distributions summed for each electronic shell. In addition, a large number of diamagnetic susceptibilities have been determined empirically. There is in general quite good agreement between the measured values and those calculated theoretically.

Molecular diamagnetism. Estimates of the diamagnetism of organic compounds are based primarily on empirical methods. From measurements on a large number of compounds, B. Pascal concluded that diamagnetic susceptibilities could be represented by the expression:

$$\chi_M = \Sigma n_A \chi_A + \lambda$$

where n_A is the number of atoms of susceptibility χ_A in the molecule, and λ is a constitutive correction depending on the nature of the bonds between the atoms. In this expression, χ_A is not the theoretical atomic susceptibility referred to in the previous section, but is a purely empirical constant derived from the measured susceptibilities. This procedure, when applied to organic compounds, often gives results that are within 1% of the experimentally determined values.

To illustrate the Pascal method, a simple example will suffice. According to this method, the molar susceptibility of ethyl bromide (C_2H_5Br) is given by $2\chi_C + 5\chi_H + \chi_{Br} + \lambda$. In this case, λ is the constitutive correction for the C-Br band. The magnitude of these quantities is as follows:

$$\{-[(2 \times 6.00) + (5 \times 2.93) + 30.6] + 4.1\} \times 10^{-6} = -53.1 \times 10^{-6}.$$

The experimentally observed molar susceptibility is -53.3×10^{-6}.

The magnetic susceptibility of a noncubic substance varies along different crystal axes. Susceptibilities measured along different crystal axes are called principal susceptibilities, and the difference between principal susceptibilities is called the magnetic anisotropy of the substance. A large amount of structural information can be obtained from measurements of principal susceptibilities. For example, the principal susceptibility of graphite perpendicular to the hexagonal axis of a single graphite crystal is about -0.5×10^{-6}, or nearly the same as the powder susceptibility of diamond. However, along the hexagonal axis, the susceptibility of graphite is -21.5×10^{-6} at room temperature. It is thought that the large diamagnetic susceptibility along the hexagonal axis is a result of the diamagnetism of conduction electrons, which are present in graphite but not in diamond.

Atomic paramagnetism. As stated previously, an atom with unpaired electrons has a permanent magnetic moment and is therefore paramagnetic. The present discussion will restrict itself to the paramagnetism exhibited by transition element compounds of the iron group and the lanthanide and actinide series. The magnetic properties of the palladium and platinum group compounds and the magnetochemistry of coordination compounds are not discussed here.

The modern quantum mechanical theory of magnetism was, to a great extent, developed by J. H. Van Vleck. One of the triumphs of this theory is the remarkable agreement between the theoretically calculated magnetic moments of the lanthanide ions and those experimentally determined. The paramagnetism of the lanthanides arises from unpaired electrons in the 4f shell which are unique because they are but little affected by the electric fields of the surrounding anions.

It is, therefore, a good first approximation near room temperature to calculate the magnetic moments of the lanthanide group ions on the assumption that their behavior is that of free gaseous ions.

A somewhat similar situation obtains with respect to the actinide series of compounds. Here, the paramagnetism arises from unpaired electrons in the incomplete 5f shell. These electrons are less well shielded from the crystalline electric fields than are the 4f electrons. Therefore, in calculating the magnetic moments of actinide ions one must take account of the splitting of the energy levels by the crystalline field. When this is done, very satisfactory agreement between calculated and observed moments is obtained in most cases.

The effect of the crystalline electric fields on the magnetic properties of ions is even more striking in the case of the iron group compounds with their unpaired 3d electrons. In general, the magnetic moment of an unpaired electron is proportional to the vector sum of the orbital and the spin angular momentum vectors. The electric fields in a crystal or a solution containing an iron group ion interact so strongly with the orbital part of the moment that they quench its contribution to the magnetic moment almost entirely. Therefore, the observed moments for these compounds are often in very close agreement with moments calculated using the so-called "spin-only" formula.

Molecular paramagnetism. Oxygen has two unpaired electrons in its normal state and is therefore paramagnetic. The molar susceptibility of oxygen over a wide range of pressures and temperatures is given by the simple equation $\chi_M = 0.993/T$ where T is the absolute temperature. Other paramagnetic gases are NO, NO_2, ClO_2, and ClO_3.

There are a large number of organic compounds which possess one or two unpaired electrons. These compounds are known as free radicals. One of the most famous examples is hexaphenylethane which dissociates in benzene solution to give the free radical, triphenylmethyl. Other compounds, such as α,α-diphenyl-β-picrylhydrazyl are stable free radicals even in the solid state.

Many organic compounds, of which fluorescein and naphthalene are examples, are normally dia- [illegible] These materials, however, become para-

magnetic when exposed to ultraviolet light. The reason for this is that light excites the molecules to triplet or phosphorescent states which are characterized by two unpaired electrons. Magnetic susceptibility measurements on such materials during irradiation have yielded important information on the mechanism of phosphorescence. Recently, paramagnetic resonance measurements on naphthalene have substantiated the earlier static susceptibility measurements.

Ferro- and antiferromagnetism. Ferromagnetic substances are distinguished chiefly by their large susceptibilities at low magnetic fields and by the fact that their specific magnetization is a function of field up to the field at which the substance is said to be saturated Above a certain temperature known as the Curie point, all ferromagnetic substances lose their ferromagnetism and become paramagnetic.

Antiferromagnetic substances also undergo a transition at a temperature which is characteristic for each material. Above this temperature, known as the Néel point, the substance is paramagnetic. The susceptibility of the material in the antiferromagnetic state is field dependent and is smaller than the susceptibility above the Néel point.

Some examples of ferromagnetic materials are iron, cobalt, nickel, gadolinium, uranium hydride, and nickel disulfide.

Examples of antiferromagnetic substances are manganese, titanium trichloride, uranium trichloride, and neptunium dioxide. *See* ATOMIC STRUCTURE AND SPECTRA; ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY; MAGNETIC RESONANCE; MAGNETISM; MOLECULAR STRUCTURE AND SPECTRA. [D.M.GR.]

Bibliography: L. F. Bates, *Modern Magnetism*, 3d ed., 1951; R. M. Bozorth, *Ferromagnetism*, 1951; P. W. Selwood, *Magnetochemistry*, 2d ed., 1956; E. C. Stoner, *Magnetism and Matter*, 1934; J. H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities*, 1932.

Magnetogas dynamics

The science of motion in a plasma under the influence of mechanical, electric, and magnetic forces; also termed hydromagnetics and magnetohydrodynamics. A plasma may be a partially or fully ionized gas, or a mixture of elementary particles such as protons, electrons, and neutrons. A plasma gas usually contains positive ions, electrons, and—if partially ionized—neutral atoms. When a gas is ionized, positive and negative charges are produced in equal numbers [illegible] ing it an electrical conductor. As such, its dynamic behavior responds to the presence of electric and magnetic fields. The effect of mechanical forces (such as those caused by a pressure gradient or by thermodynamic expansion) depends upon gas density and temperature, degree of ionization, and the strength of electric and magnetic fields.

Although magnetofluid dynamics includes the motion of any deformable, electrically conductive substance in the presence of electric and magnetic fields, the term is frequently applied specifically to liquids such as mercury. *See* ELECTROMAGNETIC PUMPS; PUMPING MACHINERY.

Interactions of particles and fields. In magnetogas dynamics, individual particle motion, as well as the macroscopic dynamics of a plasma, must be considered. An electric field transfers energy to a charged particle, accelerating it parallel to the field toward the oppositely charged electrode (Fig. 1).

A magnetic field exerts a force only if the charged particle is in motion relative to the field. Then the magnetic force is proportional to the field strength and to the velocity component which is normal to the field direction (field lines). The direction of force is normal to the field and normal to the velocity component perpendicular to the field, thereby deflecting the path of the particle (Fig. 2). In extreme cases its direction of motion may be reversed, the effect being termed a magnetic mirror (Fig. 3*a*). If the particle velocity is constant and normal to the field and if the field is constant in space and time, the particle will be trapped in the field, because continuous deflection will cause it to move in a circular path, the action being that of gyration (Fig. 3*b*). If the velocity of the particle is inclined to the field, the normal component will determine the radius of gyration, while the parallel velocity component is unaffected. Consequently, the particle describes a helical path (Fig. 3*c*). In contrast to the electrical field, the magnetic field, therefore, can merely change the direction of motion of a particle, not its kinetic energy. In the presence of crossed electric and magnetic fields,

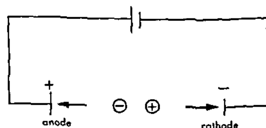


Fig. 1. Particles accelerated in an electric field.

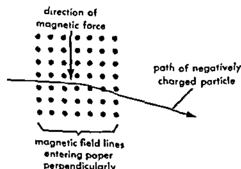


Fig. 2. Particle deflected in a magnetic field.

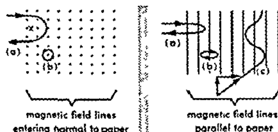


Fig. 3. Particle (a) reflected by a magnetic field; (b) trapped in a magnetic field; (c) in helical motion in a magnetic field.

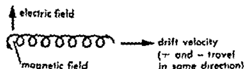


Fig. 4. Drift of particles in crossed electric and magnetic fields.

the particle is simultaneously accelerated and deflected, which causes it to drift in a direction normal to the magnetic as well as to the electric field. *See* MAGNETOHYDRODYNAMICS.

Interaction of whole plasma with fields. The two most important macroscopic aspects of magnetogas dynamics are the effect of electric or magnetic fields on the magnitude and direction of plasma motion. The use of crossed electric and magnetic fields in the manner shown in Fig. 4 causes high drift velocities. However, the macroscopic plasma motion tends to be more complex. In steady fields, complications arise because of collisions, neutralization, and re-ionization, which destroy the field pattern and dissipate energy in random velocities. In principle, these conditions can be avoided by the use of nonsteady electric and magnetic fields, such as pulsed electric and magnetic fields. In this manner, currents and fields of much higher peak intensity can be attained, resulting in more powerful electromotive forces and less time for energy dissipation in random velocities. *See* ELECTROMAGNETIC PROPULSION.

The macroscopic manifestation of charged-particle behavior in a magnetic field is that, being a conductor, the plasma is diamagnetic in that it resists penetration by a magnetic field. Field lines running parallel to the direction of plasma motion lend rigidity to the flow by resisting lateral (transverse) motion of the charges. By changing the field intensity in flow direction, the magnetic field lines converge (increasing intensity) or diverge, causing compression or expansion of the gas as in a convergent or divergent duct. Producing highly ionized plasma in an intense magnetic field, the hot gas, except for the neutral particles, can be kept away from solid walls (magnetic containment). A highly or completely ionized plasma flow represents a strong current, inducing a powerful surrounding magnetic field by means of which the plasma contains itself (pinch effect). In weaker plasma streams, the induced field has to be augmented by a field of external origin, induced by a current flowing through coils wrapped around the plasma-carrying duct. By closer coil winding, the field intensity is increased; conversely, the intensity can be decreased, corresponding to a convergent-divergent duct.

Magneto-aerodynamics. High velocity vehicles ionize the air by causing intense shock waves. This occurs at flight velocities upwards of about 18,000 ft/sec, and is experienced mainly by returning ballistic vehicles. Possibly by seeding the air in front of the vehicle with more easily ionized material (alkali metal vapor) magneto-aerodynamic conditions might be attainable down to about 9000 ft/sec. These conditions allow control of the plasma flow outside the vehicle by a magnetic field. In this manner electromotive forces can be produced which, if applied symmetrically, would cause a desirable increase in vehicle drag, and if applied unsymmetrically, would control the vehicle's attitude and, to a degree, its flight path. The drag increase due to magnetic retardation of the environmental plasma flow (electromagnetic drag) increases the effective bluntness of the vehicle, causing the shock wave to be displaced forward and to become steeper (increased shock drag). Attitude control and maneuverability without the use of moving control surfaces is an attractive possibility under the existing severe heating conditions. *See* NOSE CONE; REENTRY. [K.A.E.]

Bibliography: H. Alfvén, *Cosmical Electrodynamics*, 1950; H. W. Batten, H. L. Smith, and H. C. Early, Plasma fluctuations in crossed electric and magnetic fields, *J. Franklin Inst.*, 626(1), 1956; T. G. Cowling, *Magnetohydrodynamics*, 1957; L Spitzer, *Physics of Fully Ionized Gases*, 1956.

Magnetohydrodynamics

The science that deals with the dynamics or motion of a fluid interacting with a magnetic field. The fluid must be a good conductor of electricity and hence can be a liquid metal or (more usually) an ionized gas (or plasma). Magnetohydrodynamics is important in the development of controlled thermonuclear reactors. In these devices, the fusion reaction takes place in a high-temperature plasma composed of heavy hydrogen isotopes; the plasma is surrounded by a magnetic field which serves to confine the plasma and isolate it from the walls of the reaction chamber. Magnetohydrodynamics is employed to study the usefulness of different plasma and magnetic field configurations for this purpose *See* PLASMA PHYSICS.

Other applications include simulation of hypersonic flight conditions, ionic thrust for outer-space propulsion, space-vehicle braking upon reentry to the atmosphere, high-energy particle accelerators, microwave generators, thermionic energy-conversion devices, application of thin metallic coatings, and the study of cosmic and upper atmospheric phenomena. *See* COSMIC ELECTRODYNAMICS; METAL COATINGS.

Magnetohydrodynamics is alternately called hydromagnetics or magnetogas dynamics.

The conducting fluid and magnetic field interact [illegible] The [illegible] and [illegible] In turn, the currents influence both the magnetic field and the motion of the fluid. Qualitatively, the magnetohydrodynamic interactions tend to link the fluid and the field lines so as to make them move together.

The generation of the currents and their subse- [illegible] by the familiar laws of [illegible] drop of electric field at right angles to [illegible] tion of the motion and the field lines; the induced voltage drop causes a current to flow as in the armature of a generator.

The currents surround themselves with magnetic field lines, heat the conductor, and give rise to mechanical ponderomotive forces when flowing across a magnetic field (These are the forces which cause the armature of an electric motor to turn.) In a fluid, the ponderomotive forces combine with the pressure forces to determine the fluid motion. *See* ELECTRICITY; GENERATOR, ELECTRIC; MAGNETISM; MOTOR, ELECTRIC.

It has been suggested that magnetic disturbances which develop in the very hot and turbulent center of the sun whip up to the surface along magnetic lines of force in this way to produce sunspots. Other theories relate the heating of the solar corona and the acceleration of cosmic rays to Alfvén waves. It has been possible to generate these waves in the laboratory by twisting a column of liquid sodium placed in a strong magnetic field.

Although some early work in magnetohydrodynamics has been concerned with liquid metals, a much wider interest has developed in recent years in phenomena which involve ionized gases or plasmas. A large electrical conductivity in a plasma requires a high density of high-energy electrons. This can occur for a plasma either in thermal equilibrium at relatively high temperatures, from a few electron volts upwards (1 electron volt is equivalent to a temperature of 11,400°K), or in a non-equilibrium situation where the ions and molecules remain at a low temperature and the electrons are supplied with energy by an external source such as a microwave generator or ultraviolet radiation.

Plasmas are encountered in interstellar space, in hot stars, and in the upper atmosphere, as well as in man-made devices into which energy is fed from electrical, chemical, or nuclear sources. Strong shock waves forming ahead of a blunt object traveling with hypersonic velocities through low-density air may heat the air sufficiently to ionize it.

FUNDAMENTAL LAWS

Magnetohydrodynamic phenomena involve two well known branches of physics—electrodynamics and hydrodynamics—with some modifications to account for their interplay. *See* ELECTRODYNAMICS; HYDRODYNAMICS.

The basic laws of electrodynamics as formulated by J. C. Maxwell apply without any change. However, Ohm's law, which relates the current flow to the induced voltage, has to be modified.

It is useful to consider first the extreme case of a fluid with a very large electrical conductivity σ. Maxwell's equations predict, according to H. Alfvén, that for a fluid of this kind the lines of the magnetic field **B** move with the material. The picture of moving lines of force is convenient but must be used with care because such a motion is not observable. It may be defined, however, in terms of observable consequences by either of the following statements: (1) a line moving with the fluid, which is initially a line of force, will remain one; (2) the magnetic flux through a closed loop moving with the fluid remains unchanged.

If the conductivity is low, this is not true and the fluid and the field lines slip across each other. This is similar to a diffusion of two gases across one another and is governed by similar mathematical laws. Numerically, the distance the magnetic field will slip through the fluid in a time t is $\delta \approx \sqrt{t/\mu\sigma}$ where μ is the magnetic permeability (a constant depending upon the magnetic properties of the fluid). The condition that the conductivity be very large can now be stated more precisely: σ should be large enough so that the distance δ for the time of interest t is small compared to the dimension of the system L.

As in ordinary hydrodynamics, the dynamics of the fluid obeys theorems expressing the conservation of mass, momentum, and energy. These theorems treat the fluid as a continuum. This is justified if the mean free path λ of the individual particles is much shorter than the distances that characterize the structure of the flow. Although this assumption does not generally hold for plasmas, one can gain much insight into magnetohydrodynamics from the continuum approximation. The ordinary laws of hydrodynamics can then easily be extended to cover the effect of magnetic and electric fields on the fluid by adding a magnetic force to the momentum-conservation equation and electric heating and work to the energy-conservation equation.

The mathematical descriptions of electrodynamic and hydrodynamic phenomena—Maxwell's equations for the electromagnetic field and the equations of ordinary fluid dynamics—both involve a set of partial differential equations.

Maxwell's equations. Maxwell's equations, written in the rationalized mks system of units, are

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \qquad (1)$$

$$\nabla \cdot \mathbf{B} = 0 \qquad (2)$$

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{j} \qquad (3)$$

$$\nabla \cdot \mathbf{D} = \rho_e \qquad (4)$$

Equations (3) and (4) lead to the conservation of charge density ρ_e

$$\frac{\partial \rho_e}{\partial t} + \nabla \cdot \mathbf{j} = 0 \tag{5}$$

where $\mathbf{j}$ is the current density.

For definitions of the symbols and operations used for vectors and tensors in these and subsequent equations *see* CALCULUS OF TENSORS; CALCULUS OF VECTORS.

The electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ are related to the electric displacement $\mathbf{D}$ and the magnetic induction $\mathbf{H}$ by the equations $\mathbf{B} = \mu \mathbf{H}$ and $\mathbf{D} = \epsilon \mathbf{E}$, where μ is the magnetic permeability and ϵ is the dielectric constant for the medium. For good electrical conductors, a very small local excess or deficiency of electrons compared to the positive charge carriers would be removed almost instantly by the resulting electric field. The charge equalization takes place in a time which is roughly the larger one of the two characteristic times $t_1 = \epsilon/\sigma$ and $t_2 = \sqrt{m\epsilon/ne^2}$ where n, m, and e are the number density, mass, and charge of the electrons. In a metallic conductor, t_1 is roughly 10^{-19} sec; in the ionosphere it drops from 10^{-9} sec at a height of 100 km (E layer) to 10^{-12} sec at 250 km (F layer). Usually t_2 is larger, and the corresponding values are 6×10^{-17} sec, 6×10^{-8} sec, and 4×10^{-8} sec.

With these high rates of charge neutralization, it is not practical to calculate the electric field from the charge density through Eq. (4). The electric field is related much more effectively to the current distribution through Ohm's law

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B} + \frac{1}{\sigma}\mathbf{j} + \frac{m}{ne^2}\frac{\partial \mathbf{j}}{\partial t} \tag{6}$$

which has been reformulated to include an electric field induced by the velocity $\mathbf{v}$ of the fluid across a magnetic field and the effect of electron inertia. The latter, however, need be retained only for very rapid oscillation; otherwise the last term in Eq (6) can be dropped.

The displacement current $\partial \mathbf{D}/\partial t$ in Eq. (3) is important only when currents can pile up electrical charges. Because of the high rate with which charges are neutralized in a good conductor, it is usually possible to drop the term. This brings about a considerable simplification of Maxwell's equations, because one can now, with the help of Ohm's law, eliminate the electric field altogether and arrive at the relation

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \frac{1}{\sigma\mu}\nabla^2\mathbf{B} \tag{7}$$

The first term on the right makes the fluid and the field lines move together; the second term makes them slip across each other. The magnetic Reynolds number $R_m = vL\sigma\mu$ is a measure of the effect the motion of the fluid has on the magnetic field. When this number is very large, the second term can be dropped, producing a still simpler equation which is the mathematical basis for Alfvén's description.

Hydrodynamic equations. The conservation equations of hydrodynamics need additional terms, to take into account the interaction of the fluid with the electromagnetic field. The mass equation can remain unchanged. In the momentum equation one must add the force density $\mathbf{j} \times \mathbf{B}$ Just as the hydrodynamic force is expressed (when viscosity effects are included) as the divergence of the pressure tensor, so the magnetic force $\mathbf{j} \times \mathbf{B}$ can be expressed as the divergence of the magnetic part of Maxwell's stress tensor. This obvious analogy has led to regarding certain components of Maxwell's tensor as a magnetic pressure

The energy equation requires the addition of a term, $\mathbf{j} \cdot \mathbf{E}$, to account for the transfer of energy from the electromagnetic field to the fluid. Thus, the three hydrodynamic conservation equations modified for magnetohydrodynamics are

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{8}$$

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla \cdot \mathbf{P} + \mathbf{j} \times \mathbf{B} \tag{9}$$

$$\rho \frac{d}{dt}\left(E + \frac{v^2}{2}\right) = -\nabla \cdot (\mathbf{P}\mathbf{v} + \mathbf{Q}) + (\mathbf{j} \cdot \mathbf{E}) \tag{10}$$

Gravitational forces can be added readily if necessary.

If the mean free path $\lambda << L$, the fluid is everywhere nearly in a state of thermal equilibrium, and one can express all state variables in terms of the density ρ, the temperature T, and the velocity $\mathbf{v}$. The components of the pressure tensor $\mathbf{P}$ and of the heat flow vector $\mathbf{Q}$ are

$$P_{ij} = [p + \tfrac{2}{3}\eta(\nabla \cdot \mathbf{v})]\,\delta_{ij} - \eta\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) \tag{11}$$

$$Q_i = -K\frac{\partial T}{\partial x_i} \tag{12}$$

where $\delta_{ij} = 1$, if $i = j$; and $\delta_{ij} = 0$, if $i \neq j$. The scalar pressure p, the viscosity η, the heat conductivity K, and the internal energy E per unit mass are functions of ρ and T which depend upon the fluid and which can be found experimentally or from kinetic theory.

Using Ohm's law in the form $\mathbf{j} = \sigma(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, one can split $(\mathbf{j} \cdot \mathbf{E}) = \mathbf{v} \cdot (\mathbf{j} \times \mathbf{B}) + j^2/\sigma$ into work done by the force $\mathbf{j} \times \mathbf{B}$ and the joule heating j^2/σ. The three conservation equations lead to the entropy equation

$$\rho T \frac{dS}{dt} = \rho\left(\frac{dE}{dt} + p\frac{d}{dt}\frac{1}{\rho}\right) = \nabla \cdot (K\nabla T) + \eta\phi + j^2/\sigma \tag{13}$$

Here the rate of viscous dissipation $\eta\phi$ is positive. In many problems the three dissipation terms on the right can be ignored and the entropy S then is conserved along a streamline.

Ohm's law. The formulation of Ohm's law given in Eq. (6) is still incomplete. Yet to be added are $\mathbf{j}$

effects due to the force $\mathbf{j} \times \mathbf{B}$ and the different rate of diffusion of electrons and ions in the presence of pressure and temperature gradients. The nature of these effects depends on the size of λ in relation to the radii $r_i = \sqrt{3m_iKT}/eB$ of the spiral paths followed by particles with charge e and speed $\sqrt{3KT/m_i}$ in a field $\mathbf{B}$.

For the first time the detailed motion of the individual charged particles in the plasma must be considered. In a magnetic field, both positively and negatively charged particles describe helixes about the field lines (Fig. 1).

If a magnetic field $\mathbf{B}$ combines with an electric field $\mathbf{E}$, the center of the spiral drifts sideways in the direction of $\mathbf{E} \times \mathbf{B}$ rather than staying on a line of force. This drift is reduced roughly by a factor $1/[1 + (r_i/\lambda)^2]$ because of collisions. If $\mathbf{B}$ is small, the ion drift practically vanishes compared to the electron drift and the latter constitutes the Hall effect. If $\mathbf{B}$ is large, the two components drift nearly alike which causes a mass flow but nearly cancels the Hall current. In a strong magnetic field the effective mean free path of a charged particle in a direction perpendicular to $\mathbf{B}$ is its radius of gyration. Thus, the diffusion current is also reduced significantly by a strong magnetic field. Two forms of a generalized Ohm's law can be used:

For small $\mathbf{B}$

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B} + \mathbf{j}/\sigma + (\mathbf{j} \times \mathbf{B} - \nabla p_e)/ne \quad (14)$$

and for large $\mathbf{B}$

$$\mathbf{E}_\parallel = \mathbf{j}_\parallel/\sigma$$

$$\mathbf{E}_\perp = -\mathbf{v} \times \mathbf{B} + \left(\mathbf{j}_\perp + \frac{3n}{4B^2} \nabla kT \times \mathbf{B}\right)/\sigma_\perp \quad (15)$$

The symbols $\parallel$ and $\perp$ indicate the direction relative to $\mathbf{B}$, p_e stands for the partial pressure of the electron gas, and the conductivity $\sigma_\perp$ is approximately $\sigma/2$. For some applications it is desirable to solve Eq. (14) for $\mathbf{j}$ and this leads to

$$\mathbf{j} = \sigma \mathbf{E}_\parallel{}^1 + \sigma_1 \mathbf{E}_\perp{}^1 + \sigma_2 \frac{\mathbf{B} \times \mathbf{E}^1}{B} \quad (16)$$

where $\mathbf{E}^1 = \mathbf{E} + \mathbf{v} \times \mathbf{B} + \nabla p_e/ne$

$$\sigma_1 = \sigma/1 + \alpha^2$$

$$\sigma_2 = \alpha\sigma$$

$$\alpha = \sigma B/ne$$

The parameter α is of the same order as λ/r_e so that Eq. (16) does not apply in the limit of large α.

The electrical conductivity σ can be found experimentally or from kinetic theory. The latter leads to the formula $\sigma = ne^2\tau/m$ where τ is the effective collision time of an electron, that is, the time in which collisions alone would bring the velocity of an electron into equilibrium with that of the surrounding ions.

A partially ionized gas can be regarded as a mixture of a completely ionized plasma and a neutral gas. One can consider separate densities ρ_p, ρ_n and velocities $\mathbf{v}_p$, $\mathbf{v}_n$ for these two components.

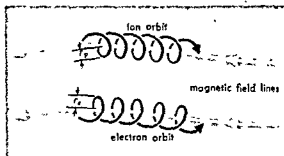


Fig. 1. Charged particles moving in a magnetic field describe helixes about field lines; sense of rotation for positive ions is opposite that for negatively charged electrons. (From *Fusion/1958, Brochure for United States Fusion Research Exhibit, Second Intern. Conf. Peaceful Uses Atomic Energy, Geneva, 1958*)

The density and velocity of the mixtures are $\rho = \rho_p + \rho_n$ and $\mathbf{v} = (\rho_p\mathbf{v}_p + \rho_n\mathbf{v}_n)/\rho$. Ohm's law retains the form given by Eq. (14) if one replaces $\mathbf{v}$ by $\mathbf{v}_p$. If one writes τ_{rs} for the effective collision time of a particle of type r with all particles of type s, the effective collision time in the formula for σ can be expressed by the relation

$$\frac{1}{\tau} = \frac{1}{\tau_{ei}} + \frac{1}{\tau_{en} + \tau_{in}}$$

The indices e, i, and n refer to electrons, ions, and neutral atoms.

The motion of the plasma relative to the neutral component is called ambipolar diffusion. It gives rise to drag forces which dissipate energy in the form of heat. The part of the dissipation resulting from the drag between ions and neutrals is most often not included in (although it is often larger than) the joule heating.

MAGNETOHYDRODYNAMIC PHENOMENA

The combination of fluid motion and electromagnetic effects can lead to much more varied phenomena than either one alone. The analysis, however, can often be simplified by observing that the equations permit similar solutions which differ only by scaling factors. The results of a particular calculation or experiment can then be applied to an entire class of arrangements. Scaling laws can be most conveniently used by specifying the value of certain dimensionless parameters such as the magnetic Reynolds number R_m mentioned before, the ratio of magnetic to kinetic energy $S = B^2/\mu\rho v^2$ and others yet to be introduced.

Equilibrium. Before turning to the peculiar interplay of hydrodynamic and electromagnetic forces, the conditions for equilibrium in the absence of motion can be considered. The formulation of equilibrium conditions can be simplified by expressing the force density as the divergence of Maxwell's stress tensor

$$(\mathbf{j} \times \mathbf{B})_i = \frac{\partial}{\partial x_k}(B_iB_k/\mu - B^2\delta_{ik}/2\mu) \quad (17)$$

The stresses represented by this tensor are twofold; a pressure $B^2/2\mu$ at right angles to the field and a tension $-B^2/2\mu$ along the field lines. The field lines thus tend to repel each other, and they also tend to contract like elastic strings. When the magnetic field lines are straight and parallel to each other, equilibrium is obtained for $p + B^2/2\mu$ = constant. When the magnetic field lines form circles around the axis of a cylinder (Fig. 2a), the equilibrium condition is

$$\frac{dp}{dr} = \frac{B}{\mu r}\frac{d}{dr}(rB) = 0$$

which requires an additional relation for its integration. For example, one can assume that B is proportional to r and this leads to $p + B^2/\mu$ = constant. The magnetic field in this geometry can balance a pressure twice as large as in the first example. Intuitively this can be understood by considering that the tension along the field lines causes an inward force adding to the magnetic pressure gradient. The particular geometry of the second example arises in the theories of the filaments in the solar atmosphere and of the static pinch in controlled fusion.

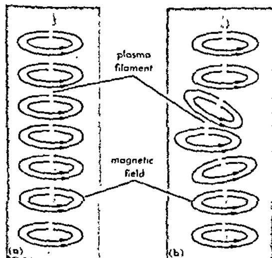


Fig 2. Plasma-magnetic-field configuration for cylindrical pinch. (a) Equilibrium configuration of plasma filament and magnetic field generated by axial current flow through plasma. (b) Onset of "kink" instability.

In discussing equilibrium configurations, it is quite common to assume simple mathematical expressions for the field. Although one usually does not consider how to generate or maintain the electric currents which give rise to the magnetic field assumed in the mathematical model, one tries to use only plausible current distributions. For instance, the magnetic field in the cylindrical geometry above could be generated by a constant current density along the axis.

It is possible to maintain appropriate combinations of toroidal and polar fields in equilibrium with both pressure and gravitational forces in axially symmetric configurations, where the fields are confined to an interior region. Specific solutions are known for the sphere and the infinite cylinder. The latter is a possible model for explaining the relatively strong magnetic fields in the spiral arms of the Galaxy.

If **B** and **j** are parallel, no force is exerted on the fluid. Such force-free fields are in equilibrium in the absence of other forces. A case of special interest is encountered if $\nabla \times \mathbf{B} = \alpha\mathbf{B}$ with constant α. Whereas normally the decay of fields due to finite conductivity creates forces which are not balanced, this special type remains force-free. An example is the twisted field in a cylinder in which the axial and tangential components of the field are proportional to the Bessel functions $J_0(\alpha r)$ and $J_1(\alpha r)$.

Stability. It is important to establish whether an equilibrium configuration is stable. To determine this, one can use an energy principle. One defines a potential energy as the sum of internal energy E, magnetic energy

$$E_m = \int (\mathbf{B}^2/2\mu)\, dv$$

and, if necessary, gravitational energy E_g. If an arbitrary deformation of the plasma always leads to an increased potential energy, the equilibrium is stable. In applying this principle, the magnetic field lines are considered to be attached to the plasma.

Another method of ascertaining stability considers small amplitude disturbances by linearizing the equations of motion. One carries out an analysis for the normal modes of oscillation and the equilibrium is unstable if any of the modes are. This analysis furnishes the rate of growth of an instability

Generally, plasma contained on the concave side of curved field lines is unstable. Important types of instability are the formation of kinks and flutes. The first is associated with cylindrical pinch discharges. A minute kink in the plasma cylinder will grow until it disrupts the discharge. Flute-shaped

change of the field and plasma between the crest and the trough. This interchange will result in a decreased potential energy and, thus, instability of the flute unless the value of $\int dl/B$ decreases toward the outside where the material pressure falls

one called the magnetic axis) do not close on themselves after a single turn around the torus. By this means it is possible to prevent interchange and thus interchange instability.

In a pinch discharge a magnetic field superimposed along the axis of the pinch stabilizes some

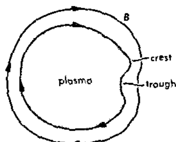


Fig. 3. Cross section of plasma cylinder with flute instability.

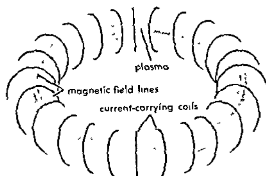


Fig. 4. Stellarator geometry. Plasma is confined in magnetic-field torus generated by external windings. Additional windings (not shown) cause field lines to rotate about chamber axis to improve stability.

but not all modes of deviation from equilibrium.

An attempt to confine the plasma on the convex side of the field has led to the cusp *geometry* (Fig 5)

A plasma configuration in equilibrium with magnetic and gravitational forces will be unstable if the magnetic field is too large. The *virial theorem* leads to an upper limit $E_m < |E_g|$ for the magnetic field energy Actually, no definitely stable configurations are known. However, in some configurations the buildup rate of instabilities is so slow they are practically stable. This is true for the cylinders which are proposed as models for galactic spiral arms.

Some interesting results appear if one enlarges the class of gravitational equilibrium *configurations* to include internal motions. In a *rotating* equilibrium configuration with axial symmetry, all points along a line of force rotate with the same angular velocity. This is known as the law of *isorotation*. The inclusion of internal motions has made possible the proof of the stability of all axis-symmetric solutions whose motion is $v = B/\sqrt{\rho\mu}$ and whose pressure p is given by $p/\rho + v^2/2 + \phi$ = constant, where ϕ is the gravitational potential.

Steady flow. A perfectly conducting fluid tends to push the field lines out of the way. With a finite conductivity, however, the field lines slip through the fluid. Generally, a large magnetic Reynolds number indicates a strong effect of the flow on the magnetic field in its path. This criterion does not depend upon the size of the magnetic field. The extent to which a magnetic field influences the motion of the fluid, on the other hand, must clearly increase with the magnitude of the field. If viscous forces are negligible in the momentum balance, the ratio of ponderomotive to inertial force $N_p = B^2L\sigma/\rho v$ is a measure of this influence.

As an example consider the one-dimensional flow of a fluid at supersonic speed coming from a field-free region and passing through a region of width L with a strong magnetic field at right angles to the direction of the flow. The magnetic field exerts a drag on the fluid and an increase in N_p causes a reduction in the rate of flow. When a critical value is passed, the flow will become subsonic. However, the fluid cannot pass continuously from supersonic to subsonic conditions. Instead, a standing shock wave will form in the magnetic region effecting the transition to subsonic flow.

In the previous example the magnetic field region was bounded by parallel planes perpendicular to the flow. Consider next an axial magnetic field confined to a cylindrical region perpendicular to the flow. In this case, a two-dimensional flow pattern will develop. The flow will be deflected sideways as it passes through the field because of [illegible]

namic applications. For strong fields the flow pattern goes around the cylinder as if it were a solid *obstacle*.

When a fluid goes around an object which has a magnetic field at right angles to its surface, it has to cross magnetic field lines. This slows the flow down and increases the size of the stagnation region (the region of zero velocity in front of the object) so that velocity and temperature gradients near the stagnation point are reduced. The magnitude of these effects also depends upon the parameter N_p where L is a typical linear dimension of the object.

Near a wall, viscous forces influence the flow and the viscosity also enters the scaling laws. The profile of a pressure-induced incompressible flow between parallel insulating walls (Poiseuille flow) with a magnetic field at right angles depends in the laminar region upon the Hartmann number $N_H = BL\sqrt{\sigma/\rho\nu}$ where $2L$ is the distance between

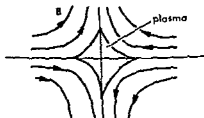


Fig. 5. Cusp geometry of plasma confined on convex side of magnetic field.

the walls (Fig. 6). For $N_H = 0$ one has the parabolic profile of classical Poiseuille flow. For large N_H the velocity is nearly constant across the channel, decreasing only in the layers adjacent to the walls. The pressure gradient necessary to maintain the same average flow velocity increases by a factor

$$N_H^2/3(N_H \coth N_H - 1) \approx 1 + N_H^2/15$$

The lines of force remain perpendicular at the walls but bulge in direction of the flow at the center line. The size of the bulge is proportional to R_m.

Frequently it is a good approximation to assume that R_m is so small that the flow does not distort the magnetic field lines, and this approximation greatly simplifies the solution of otherwise difficult flow problems. One can, for example, determine the flow around a sphere in a field which is parallel to the axis of the flow by assuming that the field is constant throughout the region of the flow. The drag on such a sphere is increased by a factor

$$1 + \frac{3}{8}N_H + \frac{7}{960}N_H^2 +$$

over the classical Stokes value.

In general, laminar flow through channels or past objects shows an increased drag with application of a magnetic field which has a component normal to the walls. The heat transfer to the wall, on the other hand, is generally reduced.

In the presence of a magnetic field a steady-state solution for the boundary layer flow along a semi-infinite flat plate can exist only when the Alfvén speed $v_a = B/(\mu\rho)^{1/2}$ is less than the speed of the flow or, equivalently, $S < 1$. This conclusion is based upon a mathematical approximation procedure carried out for the case of a longitudinal magnetic field. Either $S = 0$ or $R_m/R = \mu\nu\sigma = 0$ reduces this theory to the classical case treated by H. Blasius. If $\mu\nu\sigma$ is small but not zero the Blasius solution is not significantly modified unless S is very close to one. In the limit as $S \to 1$ the thickness of the boundary layer increases to infinity and the skin friction decreases to zero.

The case of a transverse field has been treated by an expansion in powers of $N_{px} = \sigma B^2 x/\rho v$, where x is the variable distance along the plate

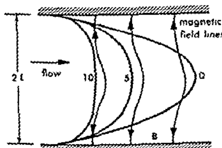


Fig. 6. Approximate profiles of flow between insulating walls for three values of N_H keeping the pressure gradient unchanged.

measured from the leading edge. This procedure does not, however, answer the question of whether there is a limit to the existence of steady flow. One can distinguish the cases in which the field is tied either to the motion of the plate or of the fluid. In the first case one finds a reduction of the skin friction coefficient at the wall of $(1 - 2.7N_{px} + 1.1N_{px}^2)$ and in the second case an increase of $(1 + 3.4N_{px} - 4.2N_{px}^2)$ compared to the Blasius solution.

Small-amplitude waves. One difficulty of the magnetohydrodynamic equations is their nonlinear character. By adopting a restriction to small-amplitude motions it is possible *to linearize* these equations. By Fourier analysis one can show the existence of a large variety of waves. Some of these are predominantly electromagnetic (with no fluid motion involved). In the presence of a dc magnetic field B_0 there are, however, low-frequency wave modes in which there is a strong interaction between fluid motion and fields. These waves are similar to ordinary sound waves, but in contrast to the ordinary fluid situation there are three sound speeds. These speeds depend strongly upon the direction of propagation relative to B_0.

The tendency of the fluid and the field lines to move together causes characteristic restoring forces which combine with the ordinary pressure forces in a number of different ways. This gives rise to the larger variety of wave types. In particular, it is possible to have shear waves with the fluid moving at right angles to the wave motion. The most simple shear wave, called the Alfvén wave, moves along the magnetic field lines with the Alfvén speed $v_a = B_0/\sqrt{\rho\mu}$. The speed of compression waves at right angles to the magnetic field is $\sqrt{v_a^2 + v_s^2}$ where v_s is the ordinary sound speed.

For an arbitrary angle θ between the direction of propagation and the direction of the magnetic field, one of the modes is a shear wave where the fluid is moving at right angles to the plane formed by the field B_0 and the wave vector k. The speed of this wave is $v_a \cos\theta$. The two other modes are in general hybrid forms with components of fluid motion both parallel and perpendicular to k and in the (B_0, k) plane. The two velocities

$$v = \sqrt{\tfrac{1}{2}[v_a^2 + v_s^2 \pm \sqrt{(v_a^2 + v_s^2)^2 - 4v_a^2 v_s^2 \cos^2\theta}]}$$

are respectively faster and slower than the shear wave velocity.

In the two extreme cases where k is parallel or perpendicular to B_0, the two hybrid modes are unscrambled into pure shear and compression modes. In the parallel case, k and B_0 do not define a plane so that the two shear modes become undistinguishable. In the perpendicular case, the velocity of both shear modes formally goes to zero and only the compression mode exists.

The strong coupling between fluid motion and field disappears as the frequency of the wave approaches $\omega_c = eB/m_i$, the angular velocity of ions spiraling in the field B. Another limit is the fre-

quency $V_a^2\mu\sigma$ at which the slipping of the fluid across the field lines destroys the wave structure.

Alfvén waves have been generated in the laboratory by twisting a column of liquid sodium in a strong magnetic field. In the design of this experiment, scaling laws have been very useful. The Lundquist number $N_L = R_m\sqrt{S} = BL\sigma\sqrt{\mu/\rho}$ seems to be natural to the description of this phenomenon. This number (where L stands for the radius of the column) must be large compared to 1, to permit the experiment. For mercury under normal laboratory conditions, say $L = 10$ cm and $B = 1000$ gauss, N_L is 10^{-1}. By using liquid sodium it is possible to increase this to $N_L = 5$. It is of interest to compare this to the interior of the earth and the sun where $N_L \sim 10^3$ and $N_L \sim 10^7$ respectively.

Shock waves. The threefold structure of magnetohydrodynamic waves that follows from the linearized theory also shows up in the nonlinear case. One is led to three speeds depending on p, ρ, and B in the same manner as before. These speeds define the characteristic motion of certain disturbances.

Just as do ordinary hydrodynamic waves, magnetohydrodynamic waves tend to get steeper in front and develop into shocks, that is, surfaces across which the physical state changes discontinuously. These changes are restricted by the conservation laws, which relate the state variables on one side of the shock front with those on the other side and with the speed u of the front. If $\Delta F = F_{\text{ahead}} - F_{\text{behind}}$ is the difference between the values of a quantity F on the two sides of the shock front and **n** is the unit vector normal to the shock front pointing ahead, the shock conditions for a fluid with infinite conductivity are

$$\Delta\mathbf{B}\cdot\mathbf{n} = 0 \quad (18)$$

$$\Delta[(\mathbf{v}\cdot\mathbf{n} - u)\mathbf{B} - (\mathbf{B}\cdot\mathbf{n})\mathbf{v}] = 0 \quad (19)$$

$$\Delta[(\mathbf{v}\cdot\mathbf{n} - u)\rho\mathbf{v} + (p + \mathbf{B}^2/2\mu)\mathbf{n} - (\mathbf{B}\cdot\mathbf{n})\mathbf{B}/\mu] = 0 \quad (20)$$

$$\Delta[(\mathbf{v}\cdot\mathbf{n} - u)\rho] = 0 \quad (21)$$

$$\Delta\left[(\mathbf{v}\cdot\mathbf{n} - u)\left(\frac{1}{2}\rho v^2 + \rho E + \mathbf{B}^2/2\mu\right) + (\mathbf{v}\cdot\mathbf{n})(p + \mathbf{B}^2/2\mu) - (\mathbf{B}\cdot\mathbf{n})(\mathbf{B}\cdot\mathbf{v})/\mu\right] = 0 \quad (22)$$

These relations can be supplemented by the statement that a discontinuity of the tangential component of **B** implies the flow of a sheet current $I = (\mathbf{n}\times\mathbf{B})/\mu$ along the front.

One can again distinguish fast, slow, and intermediate shocks. Special cases of interest are parallel and perpendicular shocks where the names indicate the direction of **B** relative to **n**. In the first of these the hydrodynamic motion is not coupled to the magnetic field and the shock proceeds as in ordinary hydrodynamics, so that the slow and intermediate types do not exist for this special case. Parallel as well as slow and fast perpendicular shocks can be considered in a frame in which the flow velocity is normal on both sides.

Shocks which are neither parallel nor perpendicular are called oblique. For these a frame of reference can be introduced in which the stream lines are parallel to **B** on both sides of the shock front but change their direction in passing through. In this frame the electric field vanishes on both sides.

In the three shock modes the change of the tangential component $\mathbf{B}_t$ of the magnetic field is distinctly different. Across a fast shock, $\mathbf{B}_t$ retains its direction and increases in magnitude; across a slow shock it retains or reverses its direction and decreases in magnitude; and across an intermediate shock it changes its direction and retains its magnitude.

In a fast shock it is possible for $\mathbf{B}_t$ to change from zero ahead of the shock front to a nonzero value behind; this is called a switch-on shock. In a slow shock it is similarly possible for $\mathbf{B}_t$ to change to zero behind the front; this is called a switch-off shock. Switch-on shocks exist only if $v_a > v_s$ ahead of the shock front and if the shock strength lies below a critical value. Switch-off shocks always exist if behind the shock front $v_a < v_s$ and they exist for $v_a > v_s$ provided the shock is strong enough.

Transient flow. Transient flows can be started either by setting an object into motion or by switching on electrical circuits which create fields.

The impulsive motion of an infinite flat plate (Rayleigh problem) in a transverse magnetic field starts a transient flow which approaches the steady state in a time of the order $\rho/\sigma B^2$.

The flow which develops along the semi-infinite flat plate in a parallel field ($S < 1$) has a different character in different regions. Between the leading edge and a point $x = (1 - \sqrt{S})vt$ the flow approaches the steady-state solution discussed earlier. Beyond $x = (1 + \sqrt{S})vt$ the flow approaches the infinite plate solution. In between there is a transition region.

The sudden release of the energy stored in a large condenser and its conversion to mechanical and thermal energy of a plasma give rise to problems whose theoretical treatment requires fast computers (*see* THERMONUCLEAR REACTION). One can simplify the theory of a collapsing pinch discharge considerably by assuming that all material which has been swept up by the contracting magnetic field is piled up in a very thin layer which is snow-plowed toward the axis. In this manner one can set up an ordinary differential equation for the radius of this layer whose numerical integration can be carried out with more modest equipment. The time of collapse of such a pinch is given by $Et \approx r(\rho\mu)^{1/4}$ where ρ and r are the initial density and radius of the plasma and E is the electric field causing the discharge.

Flow instability. Laminar flow breaks down when R rises beyond a critical number. It has been

demonstrated both theoretically and experimentally that a magnetic field improves flow stability; the critical R increases in proportion to N_H. For flow between parallel flat plates with a transverse magnetic field the predicted onset of instability takes place at $R = 50{,}000 N_H$. The measured suppression of turbulence takes place at the much lower value $R = 225 N_H$. Between coaxial cylinders, where the inner one rotates faster than the outer one, an axial magnetic field raises the critical angular velocity by a factor which approaches $N_H/2$ for $N_H \geqq 20$ (the difference in radii is used to define N_H).

Turbulence. The mathematical structure of Eq. (7) for the field vector **B** is identical with that of the equation

$$\frac{\partial \omega}{\partial t} = \nabla \times (\mathbf{v} \times \omega) + \nu \nabla^2 \omega$$

for the vorticity vector ω in ordinary hydrodynamics. The first term on the right-hand side of either equation tends to increase the mean square value of the respective field vector; the second term causes a decrease due to resistive and viscous losses. In ordinary turbulence the increase and decrease of $\overline{\omega^2}$ roughly balance each other. It has been suggested that small magnetic disturbances in a turbulent flow field will increase if resistive losses are less than viscous ones, that is, if $\sigma\mu\nu > 1$. In a Fourier expansion of the magnetic field energy, the spectrum within an uncertain range of wave numbers k tends toward a distribution $\sim k^{-5/3}$ (Kolmogoroff spectrum). Within that range the ratio of magnetic to kinetic energy is about 1.6.

Ordinary turbulence requires large values of R and magnetohydrodynamic turbulence requires even larger values of R_m. Such conditions are encountered only in geophysical and astrophysical situations.

TRANSPORT-EQUATION DESCRIPTION

The continuum approach to magnetohydrodynamics ceases to be valid when the mean free paths of the particles are of the order of or larger than the lengths characterizing the structure of the flow. A description of more general validity can be based on the functions $f_i(r,v,t)$ that give the densities of particles with velocity v at position r for the various particle species identified by the index i. These functions obey "transport equations"

$$\frac{\partial f_i}{\partial t} + v \cdot \nabla f_i + \frac{e_i}{m_i}(\mathbf{E} + v \times \mathbf{B}) \cdot \nabla_v f_i = \left(\frac{df_i}{dt}\right)_{\text{coll}}$$

where **E** and **B** are macroscopic fields that satisfy Maxwell's equations. The rate of change due to collisions $(df_i/dt)_{\text{coll}}$ can be brought into manageable form by certain assumptions. One of these is that only two particles participate in any collision. This assumption yields the Boltzmann equation. Another assumption is that collisions produce predominantly small deflections. In this case one can make certain expansions leading to the Fokker-Planck equation. Both these forms can be used to derive continuum-type equations. *See* BOLTZMANN TRANSPORT EQUATION; KINETIC THEORY OF MATTER.

Although a charged particle in a plasma collides simultaneously with very many other particles, each collision has only a minute effect. For this situation, the mean free path can be defined as the distance a particle travels until its momentum or energy is appreciably changed by the random addition of minute changes. This effective mean free path increases as the square of the temperature and can become very large. In this case one can set $(df_i/dt)_{\text{coll}} = 0$.

In a magnetic field, charged particles spiral around the lines of force; this restricts their motion at right angles to the field. If the radii of the spirals are small compared to L, these gyrations will cause some degree of randomness of the particle motion even in the absence of collisions. Gyrations are not as effective in doing this as collisions and in particular exclude any energy transfer between the components perpendicular and parallel to the magnetic field. Nevertheless, they may produce a nonthermal equilibrium in which a modified continuum approach is still possible.

One such modification replaces the scalar pressure by a tensor with two pressures $p_\parallel$ and $p_\perp$ parallel and perpendicular to **B**. With no collisions, viscous and resistive losses are absent, and therefore the assumption of no heat flow in the energy equation of the ordinary continuum theory (Eq. 13) leads to constant entropy along stream lines. In the modified theory, no heat flow leads to the constancy of $p_\perp/\rho B$ and of $p_\parallel B^2/\rho^3$ along stream lines. Such constants of motion are used mainly in investigating the stability of equilibria by means of a variational principle. However, it is not obvious why the no-heat-flow assumption should hold.

Small changes from an equilibrium distribution as they occur in applying a variational principle can also be handled by a direct use of the transport equation. This theory requires the Debye length $\sqrt{\epsilon KT/ne^2}$ to be small compared to L which is true for quite general conditions in the plasma. The stability obtained in this manner is stronger than that obtained using the constant-entropy approximation and weaker than that obtained from the two-pressure modification. *See* ELECTROMAGNETIC PUMPS; MAGNETOGASDYNAMICS; PINCH EFFECT.

[R.K.M.L.]

Bibliography: T. G. Cowling, *Magnetohydrodynamics*, 1957; R. K. M. Landshoff (ed.), *Magnetohydrodynamics*, 1957; S. Lundquist, Studies in magnetohydrodynamics, *Arkiv Fysik*, 5:297, 1952; L. Spitzer, Jr., *Physics of Fully Ionized Gases*, 1956.

Magnetometer

Primarily, an instrument for measuring the intensity of a magnetic field. As related to the earth's field, magnetometers are classed as absolute or

relative, depending on whether they are capable of calibration without direct comparison with a standard magnetic instrument. Most types of magnetometers must be standardized by measuring the intensity of a field whose strength is accurately determined by other means, thus providing a relative measurement. Among the many magnetometers in use, this article considers three of the absolute type and four of the relative type.

Absolute type. Along with the original or classical magnetometer, the sine galvanometer and the nuclear magnetometers are outstanding

The classical magnetometer. Devised by K. F Gauss in 1832 for measuring the intensity of the horizontal component of the field, this device uses a permanent bar magnet suspended in a horizontal position from a thin gold fiber. The period of oscillation in a horizontal plane depends inversely on the product of H, the intensity of the earth's horizontal field component, and M, the strength or magnetic moment of the magnet The same magnet is then used as a fixed deflector to influence the rest position of a second magnet similarly suspended from a fiber. The angle by which the second magnet is deflected from its normal undisturbed position is a measure of the ratio M/H. Thus, the complete series of observations for oscillations and deflections yields not only the value wanted, but also a check on the accuracy of the work by remeasuring the strength of the magnet, which, except for temperature effects that can be allowed for, should remain substantially constant.

The instrument is an absolute magnetometer in that it can be used for magnetic measurements without reference to other magnetic instruments; its constants are computed from accurate measurements of mass, length, time, and other physical parameters. A marked disadvantage of the universal magnetometer is that it requires from 45 min to 1 hour or more to make a complete measure-m[illegible]

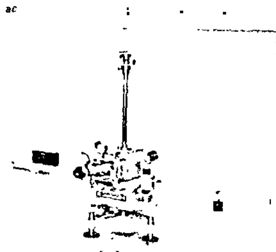


Fig 1. Observatory-type magnetometer. (U S. Coast and Geodetic Survey)

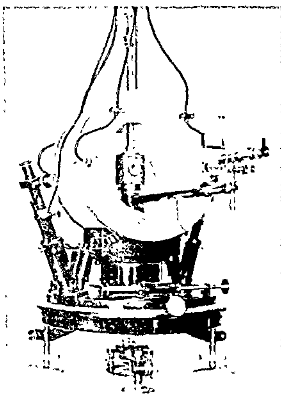


Fig. 2. Sine galvanometer. (U.S. Coast and Geodetic Survey)

lighter version for field work, the probable error of a single set is perhaps twice as large. ($1\ \gamma = 10^{-5}$ oersted.) It is used extensively in American magnetic observatories as the standard intensity instrument.

Sine galvanometer. Designed by the Department of Terrestrial Magnetism of the Carnegie Institution of Washington, this instrument employs a Helmholtz coil system wound on a hollow marble cylinder. The physical dimensions of the coils are measured to a high degree of accuracy. By sending a known current through the coils, a small magnet suspended in the coil center on a torsion-free fiber may be deflected. The rest position of the deflected magnet is a function of the intensity of the earth's field, the intensity of the coil field (computed from the coil dimensions and the current), and the angle through which the coil must be turned to preserve the alignment between magnet and coil. The accuracy of the sine galvanometer is about 0.5 γ, and a measurement can be made in about 2 min.

Nuclear magnetometers. New instruments in the class of nuclear magnetometers are rapidly coming into use because of their superiorities over older equipment. The proton-precession magnetometer, capable of measuring the total intensity (magnitude of the total field vector), requires only the accurate measurement of an audio-frequency voltage induced in a coil by the precessing protons in a cupful of ordinary water. Its absolute accuracy is limited by the uncertainty of the gyromagnetic ratio of the hydrogen nucleus, now known to about

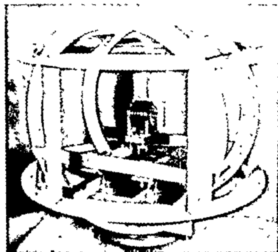


Fig. 3. Water container and Helmholtz coils of the proton vector magnetometer. (*U.S. Coast and Geodetic Survey*)

1 part in 133,000 (*see* GYROMAGNETIC RATIO). The instrument has been adapted to the measurement of the horizontal and vertical components of the earth's field (the proton vector magnetometer) through a scheme of nulling approximately the unwanted component, and measuring the remaining field by the standard method. The accuracy of the instrument is better than 0.5 γ, and a single measurement requires less than 5 sec of time.

The rubidium-vapor magnetometer, developed in 1958 by scientists of the Coast and Geodetic Survey and the U.S. National Bureau of Standards, offers prospects of high sensitivity, accuracy comparable with that of the proton-precession magnetometer, and the possibility of adaptation to measurements with space probes or satellites of very weak fields such as might be found at distances of many earth radii from the earth or in the vicinity of the moon or the nearer planets.

Relative magnetometers. Although numerous, only a few examples of relative magnetometers are considered here.

Schmidt vertical field balance. This widely used instrument consists of a bar magnet equipped with mirror and knife edges, balanced with adjustable counterpoise until it is approximately horizontal. It has been used chiefly in geophysical exploration (*see* GEOPHYSICAL EXPLORATION).

Quartz horizontal magnetometer (QHM). Developed in Denmark by D. la Cour, this apparatus is used throughout the world, both as a geomagnetic field instrument and as an observatory instrument for routine calibration of recording equipment. The instrument is remarkably simple and reliable, even though it has a rather large temperature coefficient, but it must be calibrated by observing at a standard magnetic observatory. A single one of these instruments cannot be used universally, because the size of the quartz fiber must be approximately matched with the strength of its magnet for the range of horizontal intensity in which the instrument is to be used.

Magnetometric zero balance (BMZ). Another Danish product, the so-called BMZ, is also a simple but reliable instrument for measuring differences in the vertical component. Its magnet system, constructed entirely of one piece of steel, is similar to that of the magnetic field balance. However, a turn magnet located below the balanced magnet is turned in a vertical plane until the balanced magnet reassumes its original horizontal position. Because it has been appropriately calibrated, the angle of setting of the turn magnet is then the measure of the vertical field intensity.

Saturable-core magnetometer. This electronic instrument has some advantages over older types of relative magnetometers. It has a sensing element of permalloy or similar material which becomes magnetically saturated in very low magnetic fields. A coil surrounding the core excites it to near saturation at a frequency of some 400 cps. If there is no external applied magnetic field, the alternating magnetic flux induced in the core is symmetrical in the two directions; but an external steady magnetic field along the axis of the core causes it to approach saturation more quickly during half of the cycle, and the resulting flux is asymmetric.

The saturable-core magnetometer was developed during World War II as an airborne detector of submarines. It is now used widely by both government and private organizations for geophysical prospecting surveys.

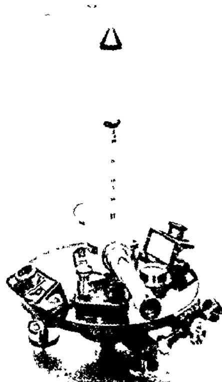


Fig. 4. Quartz horizontal magnetometer. (*U.S. Coast and Geodetic Survey*)

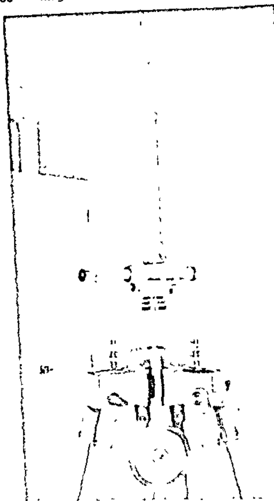


Fig. 5 Magnetometric zero balance. (U.S Coast and Geodetic Survey)

A modified form of the equipment, known as the vector airborne magnetometer, is mounted inside a large aircraft, with angle-measuring circuits connected to two servo orienting controls so that the attitude of the measuring element relative to the aircraft frame can be recorded. By providing a pendular vertical reference and a means of stellar observation for obtaining the true heading of the craft, the complete vector measurement of the magnetic field is made from an aircraft in flight. It is probable that equipment of this general type will be used as the primary instrumentation in the World Magnetic Survey proposed and sponsored by the International Union of Geodesy and Geophysics for the 5- to 10-year period immediately following the International Geophysical Year of 1957–1958. [J.H.NE.]

Bibliography: J. A. Fleming (ed.), *Terrestrial Magnetism and Electricity*, 1939; W. E. Forsythe (ed.), *Smithsonian Physical Tables*, 9th ed., 1954; D. L. Hazard, *Directions for Magnetic Measurements*, USCGS Serial 166, 3d ed., 1957; E. A. Johnson, A primary standard for measuring the earth's magnetic vector, *Terrestrial Magnetism and Atmospheric Elec.*, 44(1):29–42, 1939; H. E. McComb, *Magnetic Observatory Manual*, USCGS Spec. Publ. 283, 1952; J. H. Nelson, A new absolute instrument—the proton vector magnetometer, *J. Geophys. Research*, 63(4):880–881, 1958; E. O. Schonstedt and H. R. Irons, NOL vector airborne magnetometer type 2A, *Trans. Am. Geophys. Union*, 36(1): 25–41, 1955; T. L. Skillman and P. L. Bender, Measurement of the earth's magnetic field with a rubidium vapor magnetometer, *J. Geophys. Research*, 63(3):513–515, 1958.

Magnetomotive force

The magnetomotive force (mmf) around a magnetic circuit is the work per unit magnetic pole required to carry the pole once around the circuit. It is the analog of electromotive force. See ELECTROMOTIVE FORCE (EMF); MAGNETIC CIRCUITS.

Expressed mathematically,

$$\text{mmf} = \oint H \cos\theta \, ds$$

where $H \cos\theta$ is the component of magnetic field strength in the direction of a length of path ds. The line integral is taken around any closed path in the field.

The magnetomotive force is the rise in magnetic potential around the path. For a discussion of magnetic potential, *see* MAGNETIC FIELD.

For a path that encloses a current I,

$$\oint H \cos\theta \, ds = I$$

and for a path that encloses N equal currents, for example, a path that loops through a coil of N turns,

$$\oint H \cos\theta = NI$$

If no current is enclosed by the path, the line integral is zero.

The mks unit of magnetomotive force is the ampere-turn. *See* AMPERE-TURN. [K.V.M.]

Magneton

A unit of magnetic moment used for atomic, molecular, or nuclear magnets. *See* MAGNETIC MOMENT.

The Bohr magneton μ_B has the value of the classical magnetic moment of the electron, which can theoretically be calculated as

$$\mu_B = \mu_0 = \frac{e\hbar}{2mc}$$
$$= (0.92733 \pm .00002) \times 10^{-20} \text{ ergs/oersted}$$

where e and m are the electronic charge and mass, $\hbar$ is Planck's constant divided by 2π, and c is the velocity of light. A consistent relativistic treatment of the magnetic moment of the free electron shows that corrections to the classical calculation are necessary, so that the electron moment is about 0.1% larger than μ_0. For an extended discussion of the electron magnetic moment, *see* ELECTRON SPIN.

The magnetic moment of an atom or molecule results from contributions both from the orbital angular momentum of the atomic electrons and the electronic moments themselves (attributed to the

electron spin). When certain groups of atoms are compared, their moments show simple ratios. Observation of this fact led to the definition of the Weiss magneton (before the Bohr magneton) on a purely experimental basis as the unit for these moments. Its value is

$$\mu_W = 0.1853 \times 10^{-20} \text{ ergs/oersted}$$

The nuclear magneton is obtained from the Bohr magneton by replacing m by the proton mass, and it is thus 1836.31 times smaller than the Bohr magneton. The nuclear magneton is hardly a first approximation to the value of the nuclear magnetic moment, which is nearly three nuclear magnetons in the case of the proton (*see* NUCLEAR MOMENTS) Meson current effects are held responsible for this deviation, as well as for the fact that the neutron, which has no net charge, nevertheless has a magnetic moment of the same order (but opposite sign) as the proton. For a discussion of the neutron magnetic moment, *see* NEUTRON. [M.H.H.]

Magnetooptics

That branch of physics which deals with the influence of a magnetic field on optical phenomena Considering the fact that light is electromagnetic radiation, an interaction between light and a magnetic field would seem quite plausible It is, however, not the direct interaction of the magnetic field and light that produces the known magnetooptic effects but rather the influence of the magnetic field upon matter which is in the process of emitting or absorbing light.

Zeeman effect. This produces a splitting of spectrum lines when the emitting light source is placed in a magnetic field. The inverse Zeeman effect refers to a similar splitting of absorption lines when the absorbing substance is in a magnetic field. The Zeeman effect for spectrum lines originating from closely spaced levels is called the Paschen-Back effect (*see* PASCHEN-BACK EFFECT; ZEEMAN EFFECT). The explanation of most other magnetooptical phenomena is based on the Zeeman effect, which therefore may be regarded as the basic magnetooptic effect.

Faraday effect. This is the rotation of the plane of polarization of light when light traverses certain substances in a magnetic field. *See* FARADAY EFFECT.

Voigt effect. An anisotropic substance placed in a magnetic field becomes birefringent (doubly refracting), and its optical properties are similar to those of a uniaxial crystal. The Faraday effect is the result of this birefringence when observations are made parallel to the magnetic field. The analogous observations perpendicular to the magnetic lines of force are more difficult and were not successfully carried out until 1898 because of the smallness of the effect. The transverse magnetooptic birefringence is called the Voigt effect after its discoverer, W. Voigt.

The Voigt effect (also called magnetic double refraction) can easily be calculated for substances having a normal Zeeman effect. For more complicated Zeeman effects the results can also be theoretically predicted but are less simple quantitatively though not essentially different from those in the simpler cases.

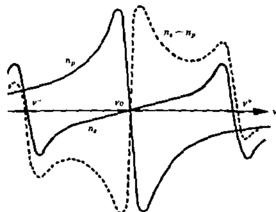
The Voigt effect depends on the fact that the indices n_s and n_p for light polarized perpendicular or parallel to the magnetic lines of force respectively are different from one another in a magnetic field where the absorption line shows a Zeeman effect. The value of n_p is independent of the magnetic field because the central component does not change while $n_s = \frac{1}{2}(n^+ + n^-)$. The dotted curve in the figure gives $n_s - n_p$, on which the observed effects of the transverse magnetic double refraction depend.

The theoretical formulas which represent the double refraction, though readily derived, are quite complicated. When the wavelength is considerably removed from the Zeeman triplet the phase difference is

$$\delta = \frac{2\pi x}{c}(n_p - n_s) = \frac{e^4 f x}{32\pi^2 c^2 n_0 (\nu - \nu_0)^3} N H^2$$

where ν_0 = absorption line frequency
ν = frequency of transmitted light
e = charge of electron
x = path length
c = velocity of light
n_0 = index of refraction without field
N = number of absorbing atoms per unit volume
H = magnetic field strength
f = so-called oscillator strength, measure of strength of absorption

Contrary to the situation existing in the Faraday effect, first order effects of magnetic double refraction are canceled out because of the presence of two perpendicular, symmetrically placed Zeeman components of combined strength equal to that of the parallel component. Because of this the Voigt effect can only be observed in the vicinity of sharply



Index of refraction of light polarized parallel (n_p) and perpendicular (n_s) to the magnetic field in the vicinity of a Zeeman triplet. The Voigt effect is proportional to $n_s - n_p$ (dotted curve).

defined absorption lines, that is, in gases and in certain crystals having sharp lines, such as rare-earth salts In the case of the rare-earth salts, a linear Voigt effect is also possible at extremely low temperatures.

Since the formula for phase difference contains the oscillator strength, f, the Voigt effect may be used to measure this important quantity.

Cotton-Mouton effect. This effect is concerned with the double refraction of light in a liquid when the liquid is placed in a transverse magnetic field. It is analogous to the electrooptical Kerr effect (*see* KERR EFFECT) and is observed in liquids with complicated molecular structure. If the molecule has a magnetic moment, the field tries to orient the molecule, but the thermal motion tends to oppose this action. There is thus a degree of orientation which depends on the temperature. If the molecule itself is optically anisotropic, the liquid also will be anisotropic and will exhibit double refraction.

The Cotton-Mouton effect is observed chiefly in nitrobenzene and aromatic organic liquids. Aliphatic compounds have a considerably smaller effect.

The phase difference of the Cotton-Mouton effect is

$$\delta = \lambda C_m x H^2$$

where x is the path length and C_m is called the Cotton-Mouton constant. For nitrobenzene at a temperature of 16.3°C and a wavelength λ of 5780 A. $C_m = 2.53 \times 10^{12}$ in absolute cgs units. With large magnets ($H = 46{,}500$ oersteds) under the most favorable circumstances, rotations of the plane of polarization up to 27° have been observed.

The dispersion of the Cotton-Mouton effect is given by Havelock's law as in the Kerr effect

Magnetooptic Kerr effect. This deals with the changes that are produced in the optical properties of a reflecting surface of a ferromagnetic substance when the substance is magnetized. In a typical case this will result in elliptically polarized light appearing in reflection, when the ordinary rules of metallic reflection would give only plane-polarized light. The component produced by the magnetic field when the magnetization is close to saturation is only of the order 10^{-3} of that normally present The explanation must be sought in the fact that the conduction electrons made to vibrate by the incident light will have a curved path in the magnetic field.

Majorana effect. This deals with optical anisotropy of colloidal solutions The effect probably is caused by the orientation of the particles in the magnetic field.

Magnetooptic effects have played an increasingly important role in microwave spectroscopy, where transitions between the Zeeman components of a single level can be observed directly (*see* MICROWAVE SPECTROSCOPY).

[C.H.D.]

Magnetoresistance

The change of electrical resistance produced in a current-carrying conductor or semiconductor on application of a magnetic field H. Magnetoresistance is one of the galvanomagnetic effects (*see* GALVANOMAGNETIC EFFECTS). It is observed both with H parallel to and transverse to the current flow. The change of resistance usually is proportional to H^2, except in very large fields, where it becomes proportional to H.

In most metals, the change of resistance is positive, however, it is generally negative in alloys of a noble metal and a transition metal and in ferromagnets above saturation.

In semiconductors, the magnetoresistance is unusually large (especially in indium antimonide) and is highly anisotropic with respect to the direction of current flow in single crystals. The latter property is of great value in determining the band structure. Magnetoresistance measurements also yield information about current carrier mobilities.

[E.A.; F.K.]

Bibliography: *See* GALVANOMAGNETIC EFFECTS.

Magnetostatics

The study of magnets and the fields produced by magnets. It should not be confused with the science of electromagnetism, which is concerned with the study of magnetic fields produced by currents. When considering magnetostatics, one looks upon magnet poles as seats of magnetism, and Coulomb's law of force between point poles in empty space is of fundamental importance. This law, expressed in rationalized mks units, and using the Sommerfeld proposal, is

$$F = \frac{\mu_0}{4\pi}\frac{m_1 m_2}{r^2} \qquad (1)$$

where $\mu_0 = 4\pi \times 10^{-7}$ weber/amp-m is the permeability of empty space; m_1 is the pole strength of one point pole and m_2 of the other, with pole strengths expressed in ampere-meters; F is the force in newtons; and r is the distance in meters between the point poles. If rationalized mks units are used but the Kennelly proposal followed, Eq. (1) becomes $F = m_1 m_2 / 4\pi\mu_0 r^2$ where μ_0 has the same value and meaning as before, m_1 and m_2 are pole strengths expressed in webers, F is in newtons, and r in meters. The force F is one of attraction if the poles are unlike and one of repulsion if they are like poles. *See* ELECTRICAL UNITS.

The magnetic field intensity H at a distance r meters in empty space from a point magnet pole of pole strength m ampere-meters is given by $H = m/4\pi r^2$ where H is in amp/meter.

For an assembly of point poles, the principle of superposition shows that the resultant H at a point P is the vector sum

$$H = \frac{1}{4\pi}\sum_{i=1}^{n}\frac{m_i}{r_i^2} \qquad (2)$$

where the south poles have minus signs and the north poles have plus signs. The direction of H at P is the direction that a north pole at P would tend to move. *See* MAGNET; MAGNETIC FIELD; MAGNETISM. [R.P.W.I.]

Bibliography: R. P. Winch, *Electricity and Magnetism*, 1955.

Magnetostriction

The change of length of a ferromagnetic substance when it is magnetized. More generally, magnetostriction is the phenomenon that the state of strain of a ferromagnetic sample depends on the direction and extent of magnetization. The phenomenon has an important application in devices known as magnetostriction transducers.

Physical cause. Magnetostriction results from the dependence of the crystalline anisotropy energy upon the state of strain of the crystalline lattice. If the crystal deforms (for example, suffers a change in length) the anisotropy energy may be lowered more than the elastic energy is raised. Thus, a strained state will be favored. For a discussion of crystalline anisotropy energy, *see* FERROMAGNETISM.

The total energy of a ferromagnetic substance depends upon the state of strain and the direction of magnetization through three contributions. The first two consist of the crystalline anisotropy energy of the unstrained lattice plus a correction which takes into account the dependence of the anisotropy energy on the state of strain. The third contribution is that of the elastic energy, which is independent of magnetization direction and is a minimum in the unstrained state. The state of strain of the crystal will be that which makes the sum of the three contributions to the energy a minimum. The result is that, when magnetized, the lattice is always distorted from the unstrained state, unless there is no anisotropy.

Since spontaneous magnetization occurs below the Curie temperature, there will always be a spontaneous lattice distortion which depends on magnetization direction in the ferromagnetic state. In nickel, the lattice spacing parallel to the magnetization is always smaller than the lattice spacing perpendicular to the magnetization.

Isotropic magnetostriction occurs when the changes in length depend only on the angle between the direction in which they are measured and the direction of the magnetization and not upon the crystal axis directions.

Magnetoelastic coupling constants. These determine the magnitude of the correction arising from strains to the anisotropy energy. In cubic crystals, there are two magnetoelastic coupling constants B_1 and B_2. The constant B_1 determines the change in anisotropy due to a diagonal component of strain, and B_2 that due to a mixed component. The values of the strain components which lead to a minimum in the total magnetoelastic energy are given in terms of the magnetoelastic coupling constants, the elastic constants, and the direction cosines of the magnetization with respect to the crystal axes, α_1, α_2, α_3. In the strain state of lowest magnetoelastic energy, that part of the energy which depends upon magnetization direction is given by

$$U = (K_1 + \Delta K)(\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2) + \cdots$$

where ΔK is a correction to the first order anisotropy constant K_1. The quantity ΔK depends only upon the magnetoelastic constants and the elastic constants.

When a high permeability (soft) magnetic material is required, the magnetostriction should be small in order that anisotropy not be induced by lattice distortions.

Applications. The magnetostrictive effect is exploited in transducers used for the reception and transmission of high-frequency sound vibrations. Nickel is often used for this application. *See* SONAR; ULTRASONICS; *see also* ELASTICITY. [E.A.; F.K.]

Bibliography: R. Becker and W. Döring, *Ferromagnetismus*, 1939; F. Seitz and D. Turnbull (eds.), *Solid State Physics*, vol. 3, 1956.

Magnetron

A microwave electron tube in which the electron beam is focused by static electric and magnetic fields which are mutually perpendicular to each other and to the direction of electron flow. Magnetrons are used as oscillators, both continuous and pulsed, power amplifiers, and voltage-tuned oscillators. Their greatest advantage over other microwave tubes is a generally higher efficiency (*see* KLYSTRON; TRAVELING-WAVE TUBE). Magnetron amplifiers are not used as low-level, sensitive amplifiers because they tend to be noisy. Magnetron-type devices are also known as M-type devices, or crossed-field devices because of the crossed electric and magnetic fields.

Magnetron oscillator. A typical magnetron oscillator is pictured in Fig. 1. The cathode is the inner cylinder, coated with an oxide or other emissive material. The anode cylinder is external and coaxial to this and contains a number of gaps across which the microwave resonant circuits are placed (*see* CAVITY RESONATOR). The useful load is coupled to one or more of the circuits. The static

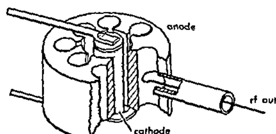


Fig. 1. Cavity magnetron oscillator.

magnetic field, usually obtained from a permanent magnet, is applied parallel to the axis of the cathode, and a radial static field results from the dc potential difference applied between cathode and anode.

Electron paths in the absence of any radio-frequency fields are illustrated in Fig. 2. Curve A, for zero magnetic field, shows the electron passing from cathode to anode under the influence of the static electric field. As the magnetic field is increased, a transverse force is produced, resulting in the curved path B. For a greater magnetic field, the path may be bent enough so that it just grazes the anode and returns to the cathode, as in C. This is the critical, or cutoff, condition for the magnetron. For still higher magnetic fields, the electron is returned to the cathode at a smaller radius, as shown by path D.

When acted on by dc fields only, all electrons returning to the cathode strike with the relatively small thermal velocity with which they initially left it, and all electrons collected by the anode arrive with the velocity v corresponding to the dc potential difference V_0

$$v = \sqrt{2(e/m)V_0} \qquad \text{meters/second} \qquad (1)$$

where e/m is electronic charge-to-mass ratio, 1.76×10^{11} coulomb/kilogram.

When ac fields from the gaps are built up, they modify the dc electron paths, adding energy to some electrons and removing energy from others. Electrons that are accelerated by the ac field follow paths of greater curvature because of the greater magnetic forces and may return to the cathode with relatively high velocities, as indicated by curve A of Fig. 3. Cathode bombardment resulting from such electrons is an important factor in cathode emission, because of its heating action and the secondary electrons produced. Electrons that are of a phase to be retarded by the ac electron fields near the gaps have less magnetic force, penetrate farther toward the anode as in path B, and are even collected as in path C. The collected electrons

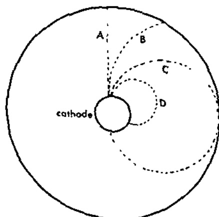


Fig. 2. Electron paths in idealized magnetron under static conditions. Magnetic field is increasing from A to D.

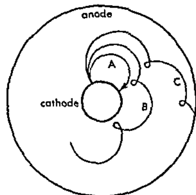


Fig 3. Representation of electron paths in magnetron with ac fields present.

arrive with a velocity less than that corresponding to the dc anode voltage. The difference in energy is delivered to the ac fields, causing these to build up the level of oscillation if the net energy flow is out of the beam into the circuit.

For energy flow to be maintained for steady-state oscillations, there must be a proper time of transit of electrons from one gap to the next in relation to the phase change of voltages from one resonant circuit to the next. There are various possible modes of the magnetron to satisfy this condition, but the most commonly used one for practical oscillators is the pi-mode. In this, the rf voltage on one gap at any instant is π radians out of phase with that on the adjacent gap, and the electrons on the average take a half cycle to rotate from one gap to the next. If other modes are close by in frequency, the coupling between these nearby modes will cause poor performance in any one mode.

Typical magnetron circuits are shown in Fig. 4 Figure 4*a* shows the hole-and-slot circuit in which the gaps contribute chiefly a lumped capacitance and the holes chiefly a lumped inductance for each resonant circuit plus an inductive coupling between holes. Strapping is normally added to separate the modes for reasons noted above. A second common circuit is the rising-sun circuit (Fig. 4*b*). Here the different lengths of alternate slots act to produce the desired mode separation. The third common circuit is the interdigital circuit (Fig. 4*c*) in which the rf voltage is developed between adjacent fingers.

Coupling to the load circuit may be by means of a loop in one cavity, as pictured in the hole-and-slot magnetron. Another common method is by means of a slit or iris coupling directly to a wave guide as pictured in Fig. 4*b*. In the interdigital circuit, two of the digits may be brought out as conductors of a two-conductor transmission system. Since the magnetron is a high-vacuum tube, a seal or window must be placed at some position in this coupling system, and for high-power applications the breakdown limit of this window is often the chief factor in limiting peak power output.

Many magnetron oscillators are built for a fixed frequency, but a reasonable range of mechanical tuning may be produced, say by introducing a series

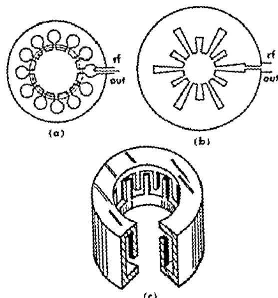


Fig. 4. Typical magnetron circuits. (a) Hole-and-slot magnetron with strapping. (b) Rising-sun magnetron. (c) Interdigital magnetron.

of rods (crown of thorns) into the holes of a hole-and slot magnetron, or by coupling strongly to an external tunable cavity. Frequency changes slightly with changes in the dc anode current (frequency pushing). Frequency changes also with variations in the reactance coupled in from the load (frequency pulling). Values of this latter effect, and also the changes in power output with changes in load, are commonly plotted on a standard transmission line chart as a Riecke diagram. An example is shown in Fig. 5.

Magnetron oscillators may be operated either continuously or pulsed. For continuous (cw) operation, powers in the kilowatt range are high, and are typically obtained with a few thousand volts. Typical pulse operation for radar purposes is with pulses of the order of a microsecond occurring approximately 1000 times a second. Peak powers over a megawatt are possible with design voltages in the tens of kilovolts. The highest-frequency magnetrons operate in the range of a few hundred kilomegacycles (wavelengths of a few millimeters) and are usually of the rising-sun type. Higher harmonics of this magnetron's fundamental frequency may also be used in obtaining useful operation at the highest frequency ranges usable by electronic means.

Magnetron amplifiers. A magnetron, M-type, or crossed-field amplifier is similar to the magnetron oscillator in that an electron stream passes near a circuit under the influence of the crossed dc electric and magnetic fields. Unlike the oscillator, however, the amplifier has the interaction circuit broken to provide an input and an output. The electron beam is normally collected also at the output of the circuit. Electrons may be introduced in a beam near the input by a crossed-field gun (Fig. 6a). The inner cylinder supporting the dc focusing field is nonemitting and is called the sole. A less common arrangement is with electrons injected by a continuously emitting sole or cathode (Fig. 6b). Magnetron amplifiers of both types may also be made linear rather than circular (Fig. 6c and d). Such arrangements are common in experimental tubes, but less so in commercial versions, because the circular arrangement requires the static focusing field over a smaller area. The interdigital circuit, of the type illustrated in Fig. 4c for magnetron oscillators, is most common for use as the slow-wave circuit in magnetron amplifiers.

Magnetron amplifiers have many of the characteristics of traveling-wave amplifiers (O-type tubes). In principle, bandwidth is as wide, but the interdigital type of magnetron circuit is easily designed with bandwidths only 10–20% of the center frequency. Gain is usually lower than for the longitudinally focused traveling-wave tube. The efficiency is appreciably higher (50–70%), because

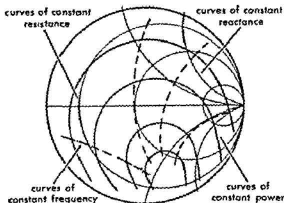


Fig 5 Riecke diagram used for showing frequency and power variations with changes in load impedance.

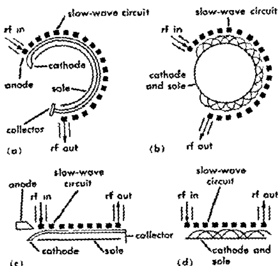


Fig. 6. Magnetron amplifiers. (a) Circular type with beam injection. (b) Circular type with continuously emitting sole. (c) Linear type with beam injection. (d) Linear type with continuously emitting sole.

the energy is taken by electrons falling through the potential drop of the crossed electric field, not by slowing down the beam as in the O-type tube. Therefore, it is easier for electron bunches to stay in step throughout the interaction in the M-type tubes. Magnetrons are most desirable for high- ... where the efficiency is a matter of

(1) less

iven de-

sired level of rf power if efficiency of conversion is higher; (2) higher efficiency also yields a lower value of dissipation on the collector and circuit, so that circuit design is easier and smaller, and simpler cooling systems are required.

One type of magnetron amplifier, called the amplitron, utilizes a backward-wave on the circuit and a beam which is allowed to continue around the tube without collection. This is similar to Fig. 6*b* but with a backward-wave circuit. This class of tubes yields high-power outputs with relatively low gains, but with high efficiencies.

Voltage tuning in magnetrons. A backward-wave oscillator may be built with M-type interaction (*see* BACKWARD-WAVE TUBE). The resulting tube, called an M-type carcinotron, or MBWO, has

commonly used for the higher-power applications. The same principle can be applied to the construction of a voltage-tuned amplifier by operating at less than the start-oscillation current, but such M-type backward-wave amplifiers were only in the experimental stage in 1959.

A second type of voltage-tuned magnetron, intrinsically much simpler, is built by designing an ordinary magnetron oscillator so that it is strongly influenced by the load. By loading down to certain values, frequency then becomes nearly a linear function of the dc voltage. Such voltage-tuned magnetrons (VTMs) are simple and lightweight, and may be used with a variety of circuits to give different products of power output and tuning range. However, they are difficult to build for the higher microwave frequencies. [J.R.WH.]

Bibliography: W. C. Brown, Description and operating characteristics of the platinotron—a new microwave tube device, *Proc. IRE*, 45:1209–1222, 1957; G. B. Collins, *Microwave Magnetrons*, 1948; J. B. Fisk, H. D. Hagstrum, and P. L. Hartman, The magnetron as a generator of centimeter waves, *Bell System Tech. J.*, 25(2):167–348, 1946; K. R. Spangenberg, *Vacuum Tubes*, 1948; H. W. Welch, Prediction of traveling-wave magnetron frequency characteristics, *Proc. IRE*, 41:1631–1653, 1953.

Magnification

A measure of the effectiveness of an optical system in enlarging or reducing an image. For an optical system that forms a real image, such a measure is the lateral magnification m, which is the ratio of the size of the image to the size of the object. If the magnification is greater than unity, it is an enlargement; if less than unity, it is a reduction.

The ratio of the longitudinal (with respect to the optical axis) dimensions of the image to the corresponding dimensions of the object is known as longitudinal magnification, which in first order equals the square of the lateral magnification.

The angular magnification γ is the ratio of the angles formed by the image and the object at the eye. The relation $n'\gamma m = n$ relates angular to lateral magnification. Here n and n' are the refractive indices of the media containing the object and image, respectively. In telescopes the angular magnification (or, better, the ratio of the tangents of the angles under which the object is seen with and without the lens, respectively), can be taken as a measure of the effectiveness of the instrument.

A small off-axis element is imaged with a different magnification in the meridional and sagittal directions. This may be called differential magnification.

Magnifying power is the measure of the effectiveness of an optical system used in connection with the eye. The magnifying power of a spectacle lens is the ratio of the tangents of the angles under which the object is seen with and without the lens, respectively. The magnifying power of a magnifier or an ocular is the ratio of the size under which an object would appear seen through the instrument at a distance of 10 in. (the arbitrary shortest distance of distinct vision) divided by the object size.

E. Abbe suggested defining the magnifying power of an optical system as the ratio of the tangent of the visual angle under which the object appears to the object size. This quantity is approximately equal to the power of the system, which is the reciprocal of the focal length, $1/f'$. *See* DIOPTER; IMAGE, OPTICAL; LENS, OPTICAL. [M.H.]

Bibliography: A. C. Hardy and F. H. Perrin, *Principles of Optics*, 1932; J. Strong, *Concepts of Classical Optics*, 1958.

Magnitude, stellar

The astronomical scale of brightness of stars and planets. A difference of five magnitudes between two objects corresponds to a factor of 100 in their rates of brightness. The basis of this system is that what the eye notes as equal intervals of brightness are actually equal ratios. One magnitude corresponds to a light ratio of $(100)^{1/5} = 2.512$.

Color and intensity. Magnitudes should be defined monochromatically, or at least over a narrow and well defined range of wavelengths; the subscript λ indicates the band used. Thus, in comparing the brightnesses I of objects A and B at a wavelength λ, the relations of magnitudes m would be

$$m_\lambda(A) = m_\lambda(B) - 2.5 \log_{10} \frac{I_\lambda(A)}{I_\lambda(B)}$$

$$\frac{I_\lambda(A)}{I_\lambda(B)} = 10^{-0.4[m_\lambda(A) - m_\lambda(B)]}$$

The energy received within two properly chosen filter bands can also be used to measure the colors of the stars. The zero point of the magnitude sys-

tem is defined so that the brightest stars are near zero; the faintest photographed (1958) with the 200-in. Hale reflector are near +23.5 magnitude. The Sun is near −27 magnitude, so that the astronomical range covers 50 magnitudes, or 10^{20} in brightness.

The flux of radiation received outside the Earth's atmosphere from the Sun over all wavelengths is 1.4×10^6 ergs/(cm²)(sec). The flux received from a star of apparent magnitude m is

$$F(m) = 2.4 \times 10^{-5}\, 10^{-0.4m} \text{ ergs/(cm}^2\text{)(sec)}$$

The energy received from a star depends on its distance r as well as on its intrinsic luminosity L. The absolute magnitude M is defined as the apparent magnitude a star would have if located at a standard distance of 10 parsecs (*see* PARSEC). From the inverse-square law, neglecting the possible interstellar absorption of light,

$$M = m + 5 - 5 \log r$$

This permits us to deduce the absolute magnitude, given apparent magnitude and distance, or the distance, given M and m. The range of M is from about −10 to +20; the absolute visual magnitude of the Sun is about +5, dependent on the wavelength interval used. The bolometric magnitude, which measures the total radiation, is about +4.6, and the luminosity of the Sun is 4×10^{33} ergs/sec Thus for any star the luminosity L is

$$L/L_\odot = 10^{-0.4(M-4.6)}$$

whence

$$L = 2.9 \times 10^{35}\, 10^{-0.4M} \text{ ergs/sec}$$

The apparent or absolute magnitude of a star depends on the wavelength range over which the measurement is carried out.

Methods of measurement. Many combinations of photographic plates, photocells, or photomultipliers (with various sensitizations), and filters (isolating various wavelength regions) are used in astronomy. Such combinations give various effective wavelengths for the centers of the response curves. Table 1 contains some of the properties of the more frequently used response bands.

The human eye and the visual magnitude system are used only for approximate work.

Color index. A wide variety of multicolor photoelectric systems are in use; correlations between such color and brightness measures are defined by the results for a group of standard stars. Although data on detector sensitivity and filter transmission partially define these systems, the actual relations found by experiment are often nonlinear.

The color of a star is given, in general, by a color index, on a particular base line

$$\text{Color index} = m_{\lambda_1} - m_{\lambda_2}$$

One frequently used index, the International Color Index, is

$$\text{IC} = m_{\text{photographic}} - m_{\text{photovisual}}$$

By convention, this and many other color syste[illegible]

Table 1. *Frequently employed magnitude systems*

Name	Detector	Filter	Response band, angstroms
Bolometric	Theoretical, sensitive to all wavelengths		
Radiometric	Vacuum thermocouple	Earth's atmospheric transmission	UV-infrared
Infrared	Photographic infrared-sensitive plate (Type IN)	Deep red filter	7000–9000
Photo red	Red-sensitive plate (103aF)	Red filter	6000–6600
Photovisual	Yellow-sensitive plate (103aD)	Yellow filter	5000–6000
Visual	*Human eye*		
Photographic	Blue-sensitive plate (103aO)	None, or minus UV	3500–5000 3900–5000
Photoelectric, V	1P21 Multiplier	Yellow	5000–6300
Photoelectric, B	1P21 Multiplier	Blue and minus UV	3900–5100
Photoelectric, U	1P21 Multiplier	Ultraviolet	3000–3900

are adjusted so that the color of bright stars of spectral type A0 is zero; such stars have color temperatures near 14,000°K. The International system is defined by the magnitudes, on these two scales, of certain stars near the north celestial pole and in other selected areas.

Many accurate photoelectric color systems exist, of which the most widely used is the Johnson-Morgan U,B,V three-color system. The accuracy of magnitudes on a photographic plate seldom exceeds ±0.1 magnitude, the precision and linearity of the photomultiplier makes ±0.01 magnitude attainable. Two color indices describe a star in the U,B,V system, that is, U-B and B-V. H. L. Johnson and W. W. Morgan (1953) have given standard colors and spectra of a large number of stars, together with the relation of the U,B,V system to the International system. Table 2 contains the mean colors as a function of spectral type. The nonlinearity of the relation between U-B and B-V colors is caused by the distortion of the energy curve of a star by continuous absorption or by lines.

Total radiation. Bolometric magnitudes m_b should measure the total energy of the star. The zero point of bolometric magnitudes is so defined that m_b

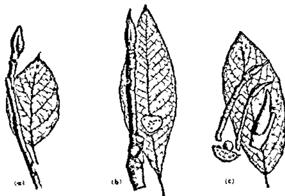
Table 2. *Magnitude of main-sequence stars in Johnson-Morgan system of standard photoelectric colors*

Type	B-V	U-B	Type	B-V	U-B
B1	−0.28	−1.00	F5	+0.44	0.00
B3	−0.20	−0.71	G0	+0.60	+0.06
B5	−0.16	−0.56	G5	+0.68	+0.21
B8	−0.09	−0.29	K0	+0.82	+0.48
A0	0.00	0.00	K3	+1.01	+0.89
A3	+0.09	+0.07	K7	+1.37	+1.2[illegible]
A7	+0.19	+0.08	M5	+1.69	+1.[illegible]
F[illegible]	+0.30	+0.02			

should equal m_v (visual) for a star of surface temperature near 6000°K, like the Sun. A bolometric correction Δm_b must be applied to m_v, negative for all other temperatures, because either the infrared or ultraviolet will contain relatively more energy at other temperatures. For cool stars Δm_b can be evaluated by radiometric techniques, the excess energy being in the accessible infrared. For hot stars, the ultraviolet, not transmitted by the Earth's atmosphere, contains most of this energy, so that Δm_b values depend on observations from high-altitude balloons or rockets or from artificial satellites. [J.L.GR.]

Magnolia

A genus of trees with large, chiefly white, flowers, and simple, entire, usually large, alternate leaves. In the winter the twigs may be recognized by their aromatic odor when bruised, by the encircling stipule scars at the nodes, and by the chambered pith.



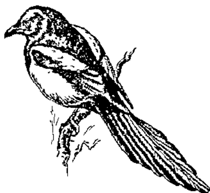
(a) Cucumber tree, *Magnolia acuminata* (b) Umbrella magnolia, *Magnolia tripetala* (c) Sweetbay, *Magnolia virginiana* (A H Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

The most important species industrially is *M. acuminata*, commonly called cucumber tree, which grows in the Appalachian and Ozark Mountains and may reach a height of 80 ft, rarely 100 ft. The fruit is red when ripe and resembles a small cucumber in shape. The wood is similar to that of the tulip tree and is rather soft, but it is of such wide natural dimensions that it is valued for furniture, cabinet work, flooring, and interior finish. The annual cut amounts to several million board feet. *See* Tulip tree.

Magnolia species occur naturally in a broad belt in the eastern United States and Central America with a similar region in eastern Asia and the Himalayas. With a few exceptions, the Asiatic species bloom early before the leaves appear. These are valued as ornamental trees in the United States. *M. stellata* is shrubby with white or pinkish flowers; *M. Yulan* is a small tree with cream-colored flowers; *M. kobus* is similar to *M. Yulan*; and *M. soulangeana*, the saucercup magnolia, has pink flowers and is supposedly a hybrid. The species of the eastern United States, all important ornamentals, bloom after the leaves appear and include *M. tripetala*, the umbrella magnolia; *M. virginiana*, the sweetbay; *M. macrophylla*, the bigleaf magnolia; and *M. grandiflora*, the southern magnolia. See Forest and forestry; Tree. [A.H.G.]

Magpie

Either of two species of birds of the family Corvidae, genus *Pica*, both of which occur in the United States. The black-billed magpie, *Pica pica*, occurs throughout the Great Basin and plains areas



The magpie, *Pica pica*; length to 21 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

of the United States eastward to Nebraska and Oklahoma, as well as in the Palearctic. The similar yellow-billed magpie, *P. nuttallii*, is limited to California west of the Sierra Nevada range. Magpies are easily recognized as black and white birds somewhat larger than jays, with long, wedge-shaped tails; they are the only land birds marked in such a manner. They sometimes live in colonies. Their principal foods are carrion, insects, and rodents. *See* Passeriformes. [J.D.B.]

Mahogany

A hard, red or yellow-brown wood which takes a high polish and is extensively used for furniture and cabinet work. The West Indies mahogany tree,



West Indies mahogany, *Swietenia mahagoni*. (USDA)

Swietenia mahagoni, a native of tropical regions in North and South America, is a large, evergreen tree with smooth pinnate leaves. Together with other species it yields the world's most valuable cabinet wood. Varieties with ornamental figures in the grain are highly prized. In the United States it occurs naturally only in the extreme southern tip of Florida, but it is planted elsewhere in the state as an ornamental and shade tree. The cigarbox or West Indian cedar, *Cedrela odorata*, belongs to the same family. *See* FOREST AND FORESTRY; TREE. [A.H.G.]

Maintainability of equipment

The inevitability of less than 100% reliability of any equipment ultimately necessitates its repair and maintenance. Also in systems of any complexity, certain trimming adjustments are necessary at various stages of assembly to maximize performance. Thus a vital consideration in the design of the components and equipments of a system is the question of how quickly and easily a unit in that system can be adjusted or repaired. In the functional design and packaging phases of component design, conveniently located test points should be provided and elements requiring adjustment and calibration should be made easily accessible. *See* SYSTEMS ENGINEERING.

Where on-site repair of components or equipments is anticipated, due consideration should be given to the ease and, in fact, possibility of replacing individual elements. The tangle of resistors, capacitors, inductances, and wires under the chassis of the typical domestic radio demonstrates the problem of part replacement. Good design dictates the regular arrangement of the elements so as to be readily accessible, removable without damage to adjacent parts, and easily and rapidly replaceable.

Where operational requirements are such that maximum availability of the equipment is essential, it is important to be able to locate defective parts rapidly and remove them easily so that replacement by operating components can be accomplished. This raises the question of proper division of the over-all system into equipments and components which can be separated from one another for fault location, removal, replacement, and repair.

A logical step beyond a policy of failed-component removal for subsequent repair is the policy of considering components to be expendable, in which case their design and packaging concepts will be planned to make the unit unrepairable. A thorough consideration of this approach leads to a concept of unitization and standardization of functional units. Complex system equipments and components are subdivided into simple, several-element, standardized building blocks. By virtue of their extreme simplicity, the design and development time of such functional units should be reduced in comparison with more conventional complex units. Since each building block represents a functional unit, existing units could be replaced as new and better versions became available. Furthermore, simple functional units are well adapted to mechanized production (*see* MINIATURIZATION OF EQUIPMENT). Assuming generalized standardization of such functional units, a wide variety of more complex component equipments and systems could be assembled by different combinations and permutations of the building blocks. The resulting high-volume manufacture of similar units might be distributed between several suppliers, broadening the procurement base for electronic equipments.

The policy of unrepairability and expendability should reduce the time and manpower involved in fault identification. With functional building blocks it is necessary to seek only the location of a functional fault, rather than an individual faulty element within the function. This allows simpler maintenance procedures, in which personnel with limited experience and ability can better become proficient.

A subdivision of equipments in functional fashion could contribute to good thermal design, or in the case where special environments are required, might facilitate the establishment of controlled environments for specific units.

Finally, the standardization of functional units, resulting as it should in higher-volume manufacture, should result in more uniform units of higher quality, therefore enhancing the reliability of the units and reliability of the over-all equipments. *See* RELIABILITY OF EQUIPMENT. [R.W.M.]

Malachite

A basic carbonate of copper with the chemical formula $Cu_2(OH)_2(CO_3)$. Malachite is normally associated with the more important copper ore deposits. Large quantities of malachite have been found in Siberia and used for ornamental stone. It has been mined as an ore of copper near Kolwezi, Belgian Congo. It is common in much smaller amounts from the western mining districts of the United States.

Malachite is monoclinic but usually occurs in massive forms or in bundles of radiating fibers. It is invariably green. It has a specific gravity of 4.05 and a hardness of 3½–4 on Mohs scale. Malachite can be readily synthesized in a number of different ways *See* COPPER. [R.I.H.]

Malacostraca

The largest and most diversified subclass of the class Crustacea which includes the shrimps, lobsters, crabs, sow bugs, beach hoppers, and their allies The shell or carapace may be large, small, vestigial, or absent; the tail or abdomen is long or short; the eyes are generally set on movable stalks but may be sessile, or even coalesced. Despite this diversity, they all share the following characteristics which demonstrate the unity of the group. The maximum number of appendages is 19 pairs, even where there is an additional body segment such as is found in the Leptostraca and embryonic Mysidacea. The trunk limbs are sharply differentiated into a thoracic series of eight pairs and an abdominal series of six pairs. The female genital duct

ways opens on the sixth thoracic segment, whereas those of the male open on the eighth.

Classification. Malacostraca are divided into two series, the Leptostraca and Eumalacostraca. An adductor muscle connects the two halves of the shell in Leptostraca. Seven abdominal segments and a pair of prongs or furcal rami on the telson are also present. Eumalacostraca lack an adductor muscle, have six abdominal segments in the adult, and (except in the order Bathynellacea) have no furcal rami. Eumalacostraca is subdivided into Syncarida, Peracarida, Pancarida, Hoplocarida, and Eucarida. The most primitive members of each subdivision approximate the common type of structure from which the more specialized members deviate widely This so-called caridoid facies (Fig. 1) is shrimplike in form, with a large carapace enveloping, but not coalescing with, the thorax. The eyes are stalked and movable, antennules are biramous, and an antennal scale is present. Swimming exopodites and respiratory epipodites are present on the thoracic limbs (Fig 2*a*) which are all similar, although there is a tendency for one or more pairs to be incorporated with the mouthparts as maxillipeds The abdomen is long, ventrally flexed with five pairs of biramous swimming appendages (pleopods), and has a tail fan formed by the lamellar uropods and the telson. These characteristics are common to Mysidacea (Peracarida), Euphausiacea and the lower Decapoda (Eucarida), and, with some modification, to the Anaspidacea (Syncarida) and Hoplocarida as well.

Evolution. The earliest recognizable fossil Eumalacostraca from Palaeozoic rocks exhibit the *caridoid facies* and some, with a brood pouch formed of overlapping plates (oostegites) arising from the coxae of the thoracic limbs (Fig. 2*c*), seem referable to the Mysidacea. From this order, series can be traced in which the carapace is gradually reduced, the thoracic exopodites are lost, and the eyes become sessile, culiminating in Cumacea, Spelaeogriphacea, Tanaidacea and Isopoda, and Amphipoda. Of the small order Thermosbaenacea (Pancarida), in which there is a dorsal carapacial

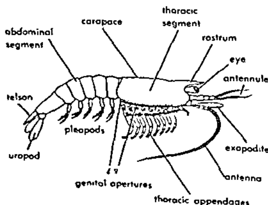


Fig. 1. Caridoid facies. (After W. T. Calman in R. E. Lankester, ed., A Treatise on Zoology, pt. 7, fasc. 3, A. and C. Black, 1909)

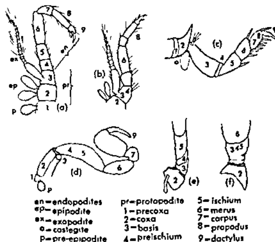


Fig. 2. Segmentation of thoracic appendages in Malacostraca. (*a*) Diagram of primitive biramous malacostracan limb, having nine segments and a terminal claw (*b*) Second limb of *Koonunga* (Syncarida) with seven segments; basis and preischium fused (*after G. Smith, 1909*). (*c*) Third limb (second pereiopod) of *Apseudes* (Peracarida), female with developing oostegite, fused carpo-propodus (*modified from H. Hansen*). (*d*) Fifth limb of *Squilla* (Hoplocarida), precoxa distinct, preischium absent or fused with ischium (*modified from H. Hansen*). (*e*) Eighth limb of *Alpheus* and (*f*) sixth limb of *Pagurus;* both showing fusion of basis, preischium and ischium (*modified from H. Hansen, 1925*).

brood pouch, no fossils are known. The other orders have no brood pouch. The Syncarida had already lost the carapace in Carboniferous times and closely resembled living Anaspidacea. The Hoplocarida, a specialized offshoot, had already assumed their characteristic structure in Jurassic times. Another series, in which the carapace coalesced dorsally with all the thoracic segments, gave rise to the Eucarida. Within the order Decapoda, by gradual reduction and flexure of the abdomen, the crablike carcinoid facies was acquired independently a number of times, resulting in mole crabs, porcelanids, king crabs, sponge crabs (*Dromia*), and the successful true crabs or Brachyura.

Terminology. The terminology applied in systematic works to the thoracic appendages and their segments in the various orders differs greatly. From one to three of the anterior pairs may be called maxillipeds and the other seven to five pairs pereiopods. The second and third (rarely also the fourth) pairs in some orders are known as gnathopods. In Stomatopoda the first five pairs are often erroneously termed maxillipeds although gnathopods would be more appropriate (Fig. 2*d*). A. J. Hansen maintains that only the first pair throughout the subclass are real maxillipeds. Although there are often seven segments (excluding the terminal claw) in a thoracic limb, the original number was apparently nine (Fig. 2*a*). These are precoxa, coxa and basis in the protopodite, and preischium, ischium, merus, carpus, propodus, and dactylus in

the endopodite. By loss or fusion of some segments, the number is frequently reduced (Fig. 2b-f). Sometimes one or more segments are secondarily subdivided. A small precoxa is distinct only in Leptostraca and a few Eumalacostraca (Fig. 2c, d). Elsewhere it is absorbed into the body wall, as is the coxa in some instances. The main flexure or "knee" joint is, according to H. Hansen, always between the merus and carpus. The exopodite arises from the basis but is frequently reduced (Fig. 2c) or absent. One or more epipodites and, in the female of Peracarida, an oostegite, may originate from the coxa; a preepipodite may occasionally be present on the precoxa (Fig. 2d). *See* EUMALACOSTRACA; LEPTOSTRACA. [I.CO.]

Bibliography: W. T. Calman, *Crustacea*, in E. R. Lankester, *A Treatise on Zoology*, pt. 7, fasc. 3, 1909.

Malaria

An infectious disease caused by the protozoan *Plasmodium*, transmitted to man by the bite of an *Anopheles* mosquito and widely distributed in endemic areas throughout the tropical and subtropical zones. It is possibly the most important single human infectious disease. Characterized by fever, enlargement of the spleen, and anemia, it has a tendency, if untreated, to recur months and sometimes years after the initial attack.

The plasmodia were first adequately described by Charles Louis Laveran in 1880, but written evidence strongly suggests that the disease occurred at least 2000 years ago in Greece, and probably in ancient China and India. In many such areas malaria is one of the great natural influences which throughout the centuries shaped the destiny of the area involved. Despite marked advances in therapy and control, made chiefly since 1930, it was estimated in 1957 that more than 200,000,000 people or about 10% of the world's population suffered from malaria and more than 2,000,000 died as a result.

The four species of plasmodia which infect man, *Plasmodium vivax*, *P. falciparum*, *P. malariae*, and *P. ovale*, all cause malaria although they provoke somewhat different symptoms. The disease is characterized by the malaria paroxysm which consists of a bout of fever accompanied first by chills, then by a sensation of excessive heat, and finally by sweating. Fever and paroxysms often recur at regular intervals with intervening fever-free periods. This clocklike regularity is distinctive and enables malaria to be distinguished from other febrile diseases, particularly when the description applies to a large population inhabiting a specific area. Appropriate chemotherapy will control the attack and may prevent recurrences, whereas untreated malaria often becomes a chronic and debilitating condition.

The diagnosis is proved by the detection of plasmodia within the red blood cells. This is usually accomplished by microscopic examination of a stained blood preparation.

A variety of different types of compounds are in use in therapy. They include quinine and other cinchona alkaloids; the 4-aminoquinoline group, chloroquine, amodiaquine, and quinacrine; the 8-aminoquinolines, pamaquin, and primaquine; a pyrimidine compound, pyrimethamine; and the biguanide, chloroguanide. Usually any one of these will cure an attack, some are more effective at preventing long-term relapses, and all provide some type of short-acting prophylaxis when taken before exposure. A nontoxic compound capable of invariably eliminating all plasmodia from the body and of giving prolonged protection when administered in a single safe prophylactic dose does not now exist.

Malaria control currently relies chiefly on insecticides, particularly in areas where funds are limited. DDT (dichloro-diphenyl-trichloroethane), with its prolonged activity, has been the mainstay in this type of control and has permitted remarkable salvage of formerly malarious areas. However, evidence of increasing insect resistance to DDT and similar insecticides is a matter of concern.

Where funds are more freely available, control based on a public-works type of program may be both effective and long lasting. These profit from the obligatory aquatic habitat of the immature stages of the mosquito, and either make the available bodies of water totally unsatisfactory for the mosquito or facilitate the application of other methods designed to render the environment hostile to anophelines. *See* DRUG RESISTANCE; HAEMOSPORIDIIDA; IMMUNITY; PARASITOLOGY, MEDICAL. [D.W.]

Bibliography: M. F. Boyd (ed.), *Malariology*, 2 vols., 1949; E. J. Pampana and P. F. Russell, Malaria: A world problem, *Chronicle World Health Organization*, 9:33–100, 1955; P. F. Russell, *Man's Mastery of Malaria*, 1955.

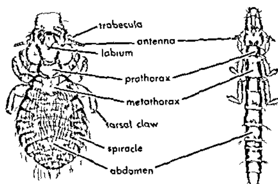
Mallophaga

A comparatively small order of insects numbering perhaps 3000 known species which are commonly called the bird lice, or more correctly the biting lice. Most of them occur among the feathers of birds. Only a comparatively small number, perhaps 300 species, occur on mammals. The group can be distinguished from the sucking lice by the fact that they always possess mandibles. Unlike sucking lice, they never have claws enlarged or modified to close about a hair or a feather. The main significance of discriminating between the biting and sucking lice lies in the fact that the Mallophaga rarely transmit any disease of their host, while the sucking lice are notorious transmitters of certain diseases. The Mallophaga may at times so increase numerically that by their sheer numbers they become an annoyance to their host, but this is the extent of their importance.

The group may be technically defined as foll relatively small insects which rarely attai in length, much flattened and wings. The antennae are 5-segmen

are distinctly developed, and their apices cross. The prothorax is developed as a distinct segment, while the mesothorax and metathorax at times are closely fused. Legs possess either one or two terminal claws; none of these ever enlarges to form grasping organs. The ovipositor is greatly reduced in size. Eggs are laid singly and are attached to the hair or feathers of the host by a drop of glue. This envelops the shaft of the hair or feather and the basal end of the egg. The young, hatching from the egg, are essentially similar to the adult. There is but slight metamorphosis. At all stages, they feed upon hair, feathers, or perhaps skin scales.

Some authorities regard the biting lice as constituting a suborder of the Anoplura. As the two groups are quite distinct, and there exists no known transition between them, they are here considered to be separate. One species is known, the louse of the elephant, which is held to constitute a suborder of the Mallophaga. It agrees with the definition in having biting mouthparts but disagrees in having the segments of the thorax completely fused. It is best treated as a distinct order.



Typical Mallophaga. (After V. L. Kellogg, Smithsonian)

Like the sucking lice, the Mallophaga are closely adapted to life upon a single host species or, at most, upon a group of closely related hosts. The physical adaptation is not as close as in the sucking lice. There is enough adaptation, however, to preclude the ready passage of a mallophagan from one host species to another, particularly at times other than in the nest or when the hosts are in close bodily contact. The result is that they are usually passed from an animal to its offspring as a sort of racial inheritance and the insects live as upon an "island" which is the host. Because of this, the classification of the Mallophaga reflects in a way the classification of their hosts.

The classification of the Mallophaga was long unsatisfactory. Consequently, the entire body of facts concerning the matter of distribution could not be synthesized. Although some aspects are evident, this question has not been given critical consideration. As the classification develops, the genera come to show, more and more closely, a restriction to special host groups.

Among the bird-infesting families, numerous examples appear of genera being restricted to a single family of birds, as *Myrsidea* to the Corvidae (crows, jays, and magpies) and *Heleonomus* to the Gruidae (cranes), although the genera of bird-infesting species have not yet been studied critically enough to demonstrate the facts.

Mallophaga

Family	Host
Boopidae	Australian marsupials
Dasyonygidae	African cavies (Procaviidae)
Gyropidae	South American rodents
Hepsogastridae	*Crypturellus tataupa* and *Nothoprocta* sp.
Laemobothridae	Aquatic birds (geese and coots)
Menoponidae	Domestic birds, sea birds
Nesiotinidae	Penguins
Philopteridae	Practically all land and water birds
Ricinidae	Numerous land and water birds
Trichophilopteridae	Lemur of Madagascar
Trimenoponidae	South American rodents

Usually no two closely related species of Mallophaga occur upon the same host. One of the strangest exceptions to this rule occurs in the family Dasyonygidae which is confined to the mammal family Procaviidae. According to some authorities, this family of mammals contains less than 10 species, but these are hosts for nearly 50 species of Mallophaga. There is no reasonable explanation for this phenomenon. See ANOPLURA; INSECTA.
[C.F.F.]

Malnutrition

A state in which there is a deficiency in one or more of the essential metabolites necessary for maintenance and growth of the intact organism. The deficiency usually arises from a dietary lack, most commonly of protein or vitamins, although deficiencies arising from a lack of essential fatty acids, minerals, and carbohydrates are occasionally seen. The severity and consequences of the malnutritional state will depend not only on the type of deficiency, but also on the length of time the deficiency exists. Most malnutritional states, however severe, are easily correctable if accurately diagnosed and treated at an early stage.

Deficiencies may also be the result of increased degradation of essential metabolites or presumably because of antimetabolites which interfere or compete with essential compounds in the cell. There apparently exists a hierarchy in the degree of importance of various cells in the body. Those that appear to be more essential include the cells controlling the activities of other cells, such as brain cells; those doing constant unexpendable work, as heart cells; certain cells in constant division, such as intestinal cells; and those producing large amounts of necessary extracellular proteins, as the liver and pancreas. These cells may be more susceptible to nutritional deficiencies than supportive cells exemplified by connective tissue cells.

Cell requirements. Most deficiencies resulting in a severe malnutritional state are complex multiple deficiencies and do not result from the loss of a single essential compound. There are seven major constituents of the cell which, in the proper con-

centrations and proper balance, are necessary in order for the cell to maintain itself. Many of these constituents have a common precursor and their pathways are closely interrelated. It is well known that the labeled carbons of radioactive glucose may eventually appear both in the carbohydrates of the cell and in the proteins, nucleic acids, and fats. The essential constituents of the cell include water, mineral ions, proteins, nucleic acids, carbohydrates, lipids, and porphyrins. Although these substances are usually found in complexes of large molecules within the cell such as the glycolipoproteins, large complex macromolecules are not transported across the cell membrane. Each individual cell is considered to be the site of synthesis of the macromolecular complexes, usually from small precursors which include amino acids, purines, pyrimidines, fatty acids, and glucose. See CELL (BIOLOGICAL).

Essential metabolite deficiencies. When considering a deficiency of the above seven essential cellular constituents, a malnutritional state resulting from protein lack would appear to be not only the most common but the most serious. Proteins make up the large portion of the solid constituents of the cell and serve not only a major structural function but also an important metabolic function, particularly when functioning as enzymes and hormones.

Proteins. Of the 20 amino acids commonly found in man, eight have been found to be indispensable. Although all 20 are required in the synthesis of most proteins, 12 of these may be manufactured by the cell itself from small carbon fragments, or other amino acids. However, if even one of the eight essential amino acids is missing from the diet, the individual is unable to synthesize any protein and growth ceases. Death does not occur immediately because many proteins are in a constant state of degradation as well as synthesis. Some unessential amino acids become available from protein breakdown for resynthesis of the more essential proteins. The essential amino acids in man are leucine, isoleucine, lysine, phenylalanine, tryptophane, threonine, methionine, and valine. Most malnutritional states do not involve an absolute deficiency of the essential amino acids but rather a relative deficiency when evaluated with the growth and maintenance requirements of the organism or cell. See AMINO ACIDS; PROTEIN METABOLISM.

Fats. With regard to the fats, only four fatty acids appear to be essential, again in the sense that only four cannot be synthesized by the cell. These are arachadonic, γ-linolenic, linolenic, and linoleic. Fats apparently serve a structural function, especially on membranes and interfaces, and usually are combined with proteins as lipoprotein. They also serve as an excellent secondary source of energy when carbohydrates are deficient or unavailable. Malnutritional states involving only fat deficiency are not well recognized. See LIPID METABOLISM.

Carbohydrates. The carbohydrates have a somewhat lesser role as structural components within the cell although their importance as a structural component in the form of extracellular polysaccharide ground substances is well known. Carbohydrates, glucose especially, are considered to be the cell's chief source of energy after degradation, oxidation, and the resultant production of high-energy adenosine triphosphate (ATP) molecules. See CARBOHYDRATE METABOLISM.

Nucleic acids and porphyrins. Nucleic acids are essential to the chromosomal structure and apparently to protein synthesis, but they may usually be synthesized from other smaller substrates. There is no evidence that malnutritional states are due to lack of nucleic acids. Some of the vitamins, however, are essential for nucleic acid synthesis and a vitamin deficiency will decrease new nucleic acid formation. Porphyrins are likewise easily synthesized in the body. See NUCLEIC ACID; PORPHYRIN.

Mineral ions. The mineral ions play a very important role in cellular metabolism. The cations sodium and potassium as well as the anions carbonate, phosphate, sulfate, and chloride are responsible for regulation of the water content of the cell and a few of them, especially potassium and magnesium, are important as activating ions for many enzymatic reactions. Malnutrition may result in an anemia due to a deficiency in iron, an ion of crucial importance in the synthesis of hemoglobin. Although sodium deficiency, hyponatremia, and potassium deficiency, hypokalemia, are seen in disease states, they are not commonly associated with malnutritional states alone. Iodine deficiency, without other evidence of malnutrition, may result in a colloid goiter of the thyroid gland. Deficiency of the other essential trace elements such as zinc, manganese, cobalt, and copper are apparently present only in severe starvation states.

Generalized malnutrition. The effect of generalized malnutrition, less than 1600 calories/day for a man weighing 70 kg, is usually obvious, resulting in extreme malaise, weakness, lack of growth, anemia, and, in severe cases, edema. The organs shrink in weight as a result of the cells shrinking in size; the fat cells in particular show characteristic changes of shrinkage and the appearance of a clear vacuolar space. Glycogen deposits in the liver and muscle disappear, and the protein structure appears reduced. Generalized malnutrition also results in a lack of resistance to any insult including drastic changes in temperature and infectious agents such as bacteria and viruses. Many of the factors concerned with this resistance are intangible. However, an important factor in resistance is the presence of antibodies. These are specialized proteins formed in response to and combined with antigenic foreign agents such as viruses. Antibody formation and thus resistance to infection is markedly low in malnutritional states.

Kwashiorkor disease. A specific disease associated with malnutrition and more specifically with a lack of dietary protein is called kwashiorkor disease. It is found most commonly in parts of A[illegible] and Asia. The patients, often children, show

vere liver disturbances. There is a reduction of the protein and nucleic acid content of the liver cells, reduction of the serum albumin, which is synthesized in the liver, and increased fat in the liver cells. Analysis of the contents of the small bowel shows markedly decreased intestinal enzyme levels. Other organs commonly affected are the muscle and pancreas which show markedly altered structure and decreased function. See DIGESTIVE SYSTEM [D.W.K.]

Bibliography: J. S. Fruton and S. Simmonds, *General Biochemistry*, 2d ed., 1958; W. C. Rose, R. L. Wixom, H. B. Lockhardt, and C. F. Lambert, The amino acid requirements of man (XV) the valine requirements: summary and final observations, *J. Biol. Chem.*, 217:987-995, 1955; R. H. S. Thompson and E. J. King (eds.), *Biochemical Disorders in Human Disease*, 1957.

Malt beverage

Fermented malt beverages are among the oldest alcoholic drinks known to mankind. Modern malt beverages include beer, ale, stout, and porter. The initial steps in the production of the various fermented malt beverages are similar.

Raw materials. Barley malt is the principal ingredient. It is not usually made in a brewery, but by the malting industry. Essentially, malt is barley which has been soaked in water and allowed to germinate, after which it is redried and ground to a flour. Malt is the source of starch-splitting and protein-splitting enzymes. The amylases hydrolyze the starch to maltose and dextrins. Maltose and some dextrins are fermentable by brewers' yeast. The proteolytic enzymes hydrolyze certain of the malt proteins to peptides and amino acids, which can be utilized by the yeast for growth. Adjuncts to malt are often used to serve as a source of starch. Common adjuncts are rice and refined corn grits, both of which are high in starch content and low in protein content. After boiling and gelatinization of these starches, they are converted to sugar by malt amylases.

The process of converting starch to fermentable sugar is called mashing and is done by various methods at temperatures of 60–70°C. The mashing methods differ in the way the temperature is raised, and this in turn affects the rate and extent of starch conversion to sugar. Finally, the enzyme action is stopped by bringing the mash to a boil, and at the same time undesirable proteins coagulate and precipitate. The extract now prepared is called the wort, and it is separated from the insoluble grain residue in a straining tank with a perforated false bottom. This tank is called a lauter tub. Sometimes filter presses are used. The hot wort is then boiled in large brew kettles, and the proper amount of hops is added to the boiling wort. Hops are the dried strobiles of the female vine *Humulus lupulus*. The hops impart the characteristic bitter flavor to beer, help in the coagulation of proteins, giving greater stability of the beer, and control to some extent the growth of many gram-positive spoilage bacteria. The wort is then strained to remove the hops, filtered to remove finer insoluble matter, cooled in heat exchangers, and aerated. Aeration favors the growth of yeast during the first part of the fermentation.

Fermentation of the wort. Inoculation of the wort with yeast is called pitching. The yeast used comes usually from an earlier brew unless laboratory inspection reveals an unusual content of contaminating bacteria. Brewery yeasts are divided into top-fermenting yeasts, which tend to rise to the top of the fermentation tank, and bottom-fermenting yeasts, which settle out during the active fermentation period. Top yeasts, usually used only for ale production, are strains of *Saccharomyces cerevisiae*, and bottom yeasts, used for the production of lager beers as well as for ale, are usually strains of *S. carlsbergensis*. The fermentation starts at about 9°C; the temperature rises a few degrees during active fermentation, and finally drops to about 5°C due to continued cooling. The exact course of a fermentation varies from plant to plant and depends also on the type of beer which is produced. The beer is allowed to rest in special tanks, during which time further impurities and yeast cells settle out. It is then filtered, sometimes with the aid of clarifying agents. Antioxidants, such as sulfite or ascorbic acid, are normally added in small amounts (maximum SO_2 content 25 ppm) since beer is subject to flavor changes on oxidation. Finally, the carbon dioxide content is adjusted to the proper level in a bottling tank, after the beer has been filtered once more. Beer packed in bottles or cans is pasteurized, but draught or keg beer is not heated.

Varieties of beer. Although the difference between some types of beer is not always clear-cut, a brief description of the main varieties is given below.

Lager beer. This is a variety of beer produced by bottom fermentation and allowed to age for some time.

Pilsner beer. This beer originated in Pilsen, Bohemia, and is light in body and color. The alcohol content is usually 3.4–3.8%. The European Pilsner beer is more heavily hopped than the American types.

Stock ale. Stock ale, or true ale, is always top-fermented and generally has a higher alcohol content than lager beer, containing up to 8% by volume. Special yeasts belonging to the genus *Brettanomyces* are sometimes used for the secondary fermentation and aging, which may take several years. These yeasts impart a slight acidity and special aroma to these ales. In addition, true ale is hopped more strongly than lager beer. In some places the word ale is used interchangeably with beer.

Porter and stout. These malt beverages are sold primarily in Great Britain and are both top-fermented ales (usually with the participation of *Brettanomyces* yeasts) of heavy body, high alcohol content, and very dark color. The color is obtained by using a dark roasted malt. Stout has a somewhat sweet taste and a strong hop character. Porter is

KEY grits water wort beer yeast fermenting wort carbon dioxide malt hops

Flow chart of brewing process. (Pabst Breweries)

also slightly sweet, but less strongly hopped than true ale.

Bock beer. Bock beer is a special brew, rather heavy, and sweeter and darker than regular beer. Whereas American beers are fairly uniform in character, many European beers show greater differences, depending on their place of origin.

Beer infections. Beers may be spoiled by wild ... and by bacteria. The bacteria are princi- ...s (*Pedio-* ...nd *Aceto-* ...NDUSTRIAL. ... M.; H.J.P.]

Bibliography: H. J. Grossman, *Grossman's Guide to Wines, Spirits and Beers*, 1943.

Maltase

An enzyme which breaks down (hydrolyzes) the disaccharide maltose into glucose. It is widely distributed throughout the animal and plant kingdoms where it occurs in association with amylase. *See* AMYLASE; ENZYME.

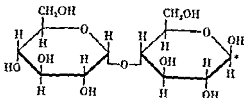
Maltase has been found in the pancreas, intestine, liver, kidney, and blood serum of animals. Among the plants it has been found in the bacteria, fungi, and monocotyledonous and dicotyledonous plants

In animals, maltase in the intestinal secretion converts the maltose (from the enzymatic hydrolysis of starch by amylase) into glucose. Glucose is then transported by the blood to the tissues for use in respiration or for conversion to glycogen by the liver. *See* DIGESTIVE SYSTEM. [D.N.L.]

Maltose

An oligosaccharide, known as malt sugar, or 4-*O*-α-D-glucopyranosyl-D-glucose, a reducing disaccharide. It crystallizes from water or aqueous alcohol as the monohydrate of the β anomer; its melting point (mp) is 102–103°C, and its optical activity is $[\alpha]_D^{20} + 112 \rightarrow + 130°$. It is fermentable by yeast in the presence of D-glucose. In the formula the asterisk denotes the reducing group.

The action of animal (salivary and pancreatic) as well as plant (germinating cereals, sweet po-



Maltose (α form; *reducing group)

tato) amylases on starch, dextrin, and glycogen produces maltose as the main end product. Maltose is hydrolyzed by acids and the enzyme maltase to

the enzymatic hydrolysis of starch and since both amylases and starch are found in the same plants, it may be a degradation product formed during the extraction process. *See* CARBOHYDRATE; GLUCOSE; OLIGOSACCHARIDE, OPTICAL ACTIVITY. [W.Z.H.]

Malvales

An order of the plant subclass Dicotyledoneae, including 5 families having 202 genera and 2915 species The mallow family contains many ornamentals such as hibiscus, Turk's cap, rose of China, and hollyhock. It also contains some species of great economic importance, such as marsh mallow, okra, and cotton. The Sterculia family (Sterculiaceae) is the source of chocolate and cola, and also contains many ornamentals. The linden family (Tiliaceae) includes the linden or basswood whose flowers supply choice nectar for bees, and the wood is used in making the beekeeper's supplies. Also included in this family is jute, a fiber obtained from asiatic species The bombax family (Bombacaceae) is an assemblage of tropical trees including the baobab, cottontree, silk cotton (source of kapok), and balsa, the lightest commercial wood in the world. *See* BALSA; BASSWOOD; COCOA POWDER AND CHOCOLATE, COLA; COTTON; JUTE; KAPOK TREE; OKRA, *see also* DICOTYLEDONEAE, EMBRYOPHYTA; PLANT KINGDOM, TREE. [P.D.S.]

Mammalia

A class of vertebrates in which the young are nourished with milk from special milk-secreting glands of the mother. Mammals are warm-blooded and are typically covered with hair, although certain forms, notably the whales and sirenians, have secondarily become almost completely hairless. The lower jaw is reduced to a single pair of bones, the dentaries, which are hinged directly to the squamosal without the intervention of a quadrate. Sound waves are transmitted across the middle ear by a chain of three bones (malleus, incus, and stapes), derived from elements of the reptilian jaw and jaw-suspension apparatus. The heart is four-chambered, and blood leaves the heart by way of a left aortic arch. About 3800 species of living mammals are known. *See* EAR; LACTATION; MAMMARY GLAND.

Phylogeny. Mammals arose from therapsid reptiles during the Late Triassic, about 175,000,000 years ago, and the transition from mammal-like reptile to true mammal is well documented in the fossil record. It is evident that the class Mammalia is polyphyletic; that is, the boundary between reptile and mammal was crossed independently by several different stocks. By the beginning of the Jurassic the mammals were represented by at least three subclasses, each representing a separate crossing of the reptile-mammal threshold: the Prototheria (*docodonts* and monotremes), Allotheria (multituberculates), and Theria (pantotheres and symmetrodonts). A group of peculiar early mammals, *the Triconodonta*, cannot be allocated with certainty to any of these subclasses, and probably represents a fourth crossing of the reptile-mammal threshold. They are omitted from the following classification. See separate articles on each group.

Class Mammalia
- Subclass Prototheria
  - Order Monotremata
- Subclass Allotheria
  - Order Multituberculata
- Subclass Theria
  - Infraclass Pantotheria
    - Order Pantotheria
    - Order Symmetrodonta
  - Infraclass Metatheria
    - Order Marsupialia
  - Infraclass Eutheria
    - Order Insectivora
    - Order Dermoptera
    - Order Chiroptera
    - *Order Primates*
    - Order Tillodontia
    - Order Taeniodonta
    - Order Edentata
    - Order Pholidota
    - Order Lagomorpha
    - Order Rodentia
    - Order Cetacea
    - Order Carnivora
    - Order Condylarthra
    - Order Litopterna
    - Order Notoungulata
    - Order Astrapotheria
    - Order Tubulidentata
    - Order Pantodonta
    - Order Dinocerata
    - Order Pyrotheria
    - Order Proboscidea
    - Order Embrithopoda
    - *Order Hyracoidea*
    - Order Sirenia
    - Order Perissodactyla
    - Order Artiodactyla

The modern marsupial and placental stocks arose from pantotheres during the Cretaceous and radiated into the modern orders of mammals during the Late Cretaceous and the Paleocene and Eocene. All marsupials (Metatheria) are placed in a single order, but the placentals (Eutheria) are divided into 26 orders, 10 of which became extinct during the Cenozoic. The largest order of Eutheria [illegible] followed

mates. *See* MAMMALIA FOSSILS; TRICONODONTA.

Morphology. Their whole morphology indicates that mammals differ more from other vertebrate classes than any of these do among themselves. The transition from reptile to mammal was, in fact, more a revolution than an evolution of the vertebrate morphological pattern. The significance of

these far-reaching changes is unknown. Birds, whose evolution in many ways parallels that of mammals, are far more reptilian, or even fishlike, than are mammals.

Geographic distribution. The present geographic distribution of mammals has a complex history. Many orders and families arose in Eurasia and North America and spread to the south. Africa south of the Sahara has been more or less isolated since early in the Tertiary, and several forms such as the golden mole, aardvark, okapi, and elephant shrews are peculiarly African, whereas other forms common elsewhere, such as deer, sheep, goats, and bears, are completely absent. South America was isolated through most of the Tertiary and developed a very peculiar fauna based on mammals that reached there before the contact with North America was broken. When the Panama land bridge was reestablished in the late Tertiary, modern mammalian types entered from the north and exterminated most of the original inhabitants. Australia and Madagascar were both completely isolated throughout the Tertiary, and have remarkable mammalian faunas. The Australian fauna is almost wholly marsupial; the Madagascan fauna features primitive insectivores, generalized lemurs, and primitive carnivores. *See* ZOOGEOGRAPHIC REGION; ZOOGEOGRAPHY. [D.D.]

Mammalia fossils

The earliest mammals, which were contemporaries of the Jurassic and Cretaceous dinosaurs, evolved from theriodont reptiles in the Late Triassic. The

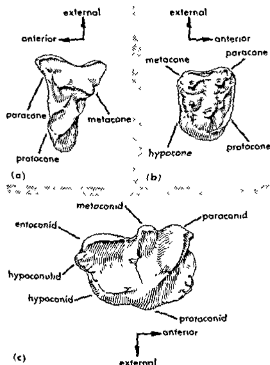


Fig. 1. Mammalian molars, principal cusps labeled (a, b) Occlusal views of upper molars. (c) Occlusal view of lower molar.

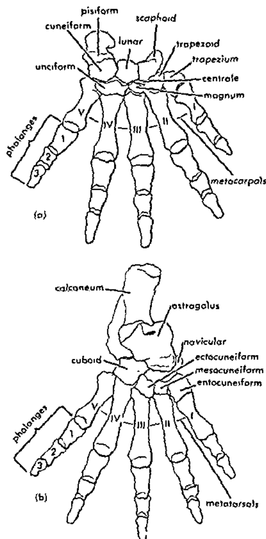


Fig. 2. Primitive mammal. (a) Anterior views of right forefoot. (b) Hind feet (After Matthew, 1937)

available fossils seem to demonstrate satisfactorily that the animals currently recognized as mammals arose from different groups of advanced mammal-like reptiles. Of the characters that define the class Mammalia, the presence of a squamosal-dentary jaw articulation has been used by paleontologists to separate primitive mammals artificially from their reptilian ancestors. Some of the other important characters of mammals are the presence of hair, a controlled body temperature, and lacteal glands. Most mammals are viviparous, have cheek teeth with two or more roots, and usually have seven cervical vertebrae. *See* THERAPSIDA.

Toward the end of Cretaceous time, after the initial radiation of the flowering plants, the small Mesozoic mammals gave rise to a great variety of large and small forms destined to be an important element in the land life of the succeeding Cenozoic Era. Mammals have a high adaptive potential which enabled some of them to adapt to the speci requirements of aquatic and volant modes of . Cenozoic mammals inhabited every landmas

cept possibly Antarctica, and marine forms reached that remote region. During the Cenozoic, mammals were being constantly modified as they dispersed and adapted to the changing conditions. Their rapid change and rapid dispersal make fossil mammals very useful for correlating the rocks in which they were preserved.

Any discussion of fossil mammals must involve a discussion of the characters of the teeth and feet, for these are the most commonly fossilized portions of the body. The terminology employed in describing the cusps of the teeth and the bones of the feet of mammals is illustrated in Figs. 1 and 2. *See* DENTITION; EVOLUTION, ORGANIC; PALEONTOLOGY. [R.A.ST.; W.A.CL.]

Bibliography: C. L. Camp et al., *Bibliography of Fossil Vertebrates*, Geol. Soc. Am. Spec. Papers and Mem., 1940, 1942, 1949, J. Piveteau, *Traité de paléontologie*, vols. 6 (pt. 1) and 7, 1957–1958, G. G. Simpson, *The Principles of Classification and a Classification of Mammals*, Bull. Am. Museum Natl. Hist., 85(16), 1945.

Mammary gland

One of the distinguishing and unique anatomical structures of Mammalia designed to secrete a nutrient for the newborn suitable to bridge the transition from intra- to extrauterine life. The most primitive mammal, the duckbill or platypus, resembles reptiles in laying eggs. Its mammary glands are unique in lacking nipples which are present in all other mammals. The milk oozes out of two mammary gland areas and is lapped up by the young.

In all other mammals, the young are born alive. In marsupials such as the kangaroo, there is a pouch associated with the mammary glands, for in these mammals the young are born in a very immature state of development. The embryo never becomes attached to the maternal organism. Following birth the young crawl from the uterus into the mammary pouch and become attached to a nipple for a considerable period. The marsupial mammals are common in Australia. The opossum is the only North American representative, but several ratlike forms live in South America. They are not found in the wild state in Europe, Asia, or Africa.

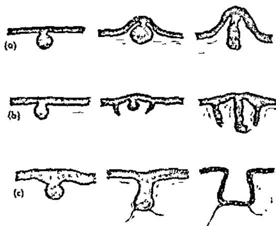


Fig. 1. Types of nipple or teat development. (a) Proliferation teat. (b) Epithelial ingrowth teat. (c) Eversion teat. (From C. W. Turner, The mammary glands, in E. Allen, ed., Sex and Internal Secretions, 2d ed., Williams and Wilkins, 1939)

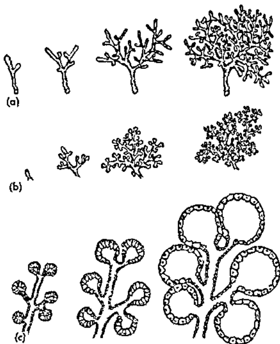


Fig. 2. Schematic diagram of development of the mammary gland. (a) The mammary gland develops from a simple branched tubular gland into a compound tubular structure during recurring estrous cycles. (b) The proliferation of the ends of the tubular ducts into a vast number of alveoli, transforming the tubular gland into a typical compound tubuloalveolar gland, occurs during the first half of pregnancy. (c) The cells of the alveoli enlarge with secretion during the last part of pregnancy. The discharge of this secretion into the lumina of the alveoli causes the enlargement of the gland observed at the approach of parturition. *(From C. W. Turner, The Mammary Glands, in E. Allen, ed., Sex and Internal Secretions, 2d ed., Williams and Wilkins, 1939)*

Placental mammals. The placental mammals represent 95% of the living mammals. The young are nourished for a considerable period in the uterus through the mediation of vascular connections of maternal and fetal membranes. This group includes man and other primates, domestic animals which furnish the milk supplies, and many experimental and wild mammals. *See* EXTRAEMBRYONIC MEMBRANES.

Embryology. The mammary glands are considered skin glands because their embryonic origin stems from the proliferation of the ectoderm which gives rise to the skin (*see* GLAND). The embryonic development includes cell multiplication in the form of two mammary lines with discrete circular proliferation in the lines in the general location of the future glands. These proliferating areas be-

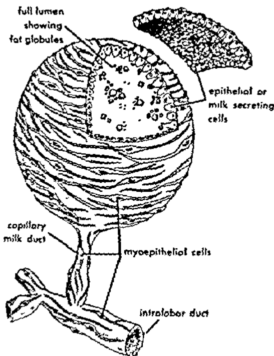


Fig. 3. An alveolus where milk is made. There are literally billions of them, microscopic in size, in a cow's udder. (*From C. W. Turner, Harvesting your milk crop, reprinted from The Surge News, winter issue, 1954–55*)

come spherical or cuboidal structures called mammary buds. From the basal end of the buds, one or more primary sprouts develop into the underlying tissue. In man there may be as many as 20 primary sprouts arising from the mammary bud. At the same time the nipple or teat begins to develop in one of three different ways. The most common type is a proliferative nipple in which the underlying tissue elevates the mammary bud as the primary sprouts grow downward. The second type is called an eversion nipple because the bud sinks into a deep pocket from which the nipple emerges after sexual maturity by everting. This type of teat development is present in many marsupials. In the third type, present in rats and mice, the nipple is formed by an epithelial ingrowth around the gland area which permits the nipple to emerge at sexual maturity.

Maturation. At birth, the mammary gland consists of the nipple and one or more ducts arising from primary and secondary sprouts. Until the approach of puberty, little further extension and sprouting of the duct system occurs. Then, through the stimulus of estrogen and recurring estrous cycles, considerable duct growth occurs to form fatty pads or breast tissue.

Effect of pregnancy. When animals become pregnant, the growth of the duct system continues with the development of many side branches. At the sides and ends of these branched ducts, a new type of spherical cellular growth occurs called alveoli. These structures may be compared to individual grapes and the bunch of grapes to a lobule, each of which is supported and surrounded by connective tissue. The alveoli are lined internally by epithelial cells capable of secreting milk into the central expandable cavity called the lumen. During the period of mammary gland lobule-alveolar growth, the lumen of the alveoli do not show because milk is not secreted.

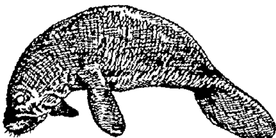
Lobule-alveolar growth is stimulated during the first two-thirds of pregnancy by the simultaneous secretion of estrogen and progesterone, first by the ovary and later by the placental membranes as well. The number of alveolar cells formed determines the potential milk-secreting capacity of the mammary glands of the individual animal. The yield of milk in many animals is limited by poor growth of the mammary gland. A hormone secreted by the anterior lobe of the pituitary, called mammogen, plays a vital role in mammary gland growth. *See* OVARY; PITUITARY GLAND.

During the final third of pregnancy, the cells of the mammary gland begin to secrete a fluid called colostrum. The cells elongate and discharge their secretion into the lumen of the alveoli. The colostrum accumulates in the glands and causes a gradual enlargement of the breast or udder. At the time of parturition the mammary glands are fully prepared to secrete. The functional activity of the glands is controlled by the lactogenic hormone secreted by the anterior pituitary. Various constituents of the blood coming to the mammary gland are transformed or synthesized into the unique constituents of milk through the mediation of enzyme systems in the cells. *See* LACTATION. [C.W.T.]

Manatee

Any of three large aquatic animals of the family Trichechidae, order Sirenia, confined to the Atlantic Ocean, with representatives in American and African waters. These are large mammals, often called sea cows, and are highly modified for aquatic life. They have no hindlimbs; the forelimbs are modified into flippers; and the posterior end of the body is flattened into a flipper with lateral flukes. The skeleton is unusually heavy. Manatees have thickened, fleshy lips and teeth somewhat similar to those of the elephant. Their nostrils have valves.

The Florida manatee, *Trichechus manatus*, is found in United States waters, frequenting river



The manatee, *Trichechus manatus*; length to 12 ft. (*From P. M. Duncan, ed., Cassell's Natural History, Cassell, 1883*)

mouths and lagoons from North Carolina around Florida to Texas and Mexico. It is nowhere common, but is more often found along the southern Florida coast than elsewhere. See SIRENIA. [J.D.B.]

Mandarin

A name used to designate a large group of citrus fruits in the species *Citrus reticulata* and some of its hybrids. This group is variable in character of trees and fruits since the term is used in a general sense to include many different forms, such as tangerines, King oranges, Temple oranges, tangelos (hybrids between grapefruit and tangerine), Satsuma oranges, and Calamondin, presumably a hybrid between a mandarin and a kumquat (*see* GRAPEFRUIT; KUMQUAT, ORANGE; TANGERINE).

Many varieties of citrus falling in the mandarin classification are compact trees with willowy twigs [illegible] The fruits have loose [illegible] ..ng feature common to the true mandarins and their hybrids is the presence of chlorophyll in the cotyledons (seed leaves), giving them a pistachio-green color.

Although tangerines are the most extensively planted of the mandarin group, certain others, particularly the Temple orange, the Murcott orange, and the tangelos, of which there are several varieties, are important commercial fruits in the United States. *See* FRUIT (BOTANY); FRUIT (TREE) [F.F.C.]

Mandibulata

A subphylum of the phylum Arthropoda. The Mandibulata, also known as the Antennata, includes the classes Crustacea, Insecta, Chilopoda, Diplopoda, Symphyla, and Pauropoda. The last four classes were included originally in the class Myriapoda *which is not* generally recognized as a natural

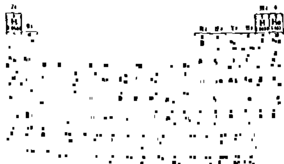
General characteristics of the Mandibulata

Characteristic	Crustacea	Insecta	Myriapoda
Tagmata or body divisions	Usually cephalothorax, abdomen	Head, thorax, abdomen	Usually head and body
Antennae	2 pairs	1 pair	1 pair
Mouth parts	1 pair mandibles, 2 pairs maxillae	1 pair mandibles, 1 pair maxillae	1 pair mandibles, 1 or 2 pairs maxillae
Legs	1 pair per somite or none	3 pairs on thorax	1 or 2 pairs per somite
Development	Usually with larval stage	Usually with larval stage	Direct

group; however, the term is still employed. The mandibulate arthropods *possess a pair of mandibles* or jaws which *characterize the* group. *See* CHILOPODA; CRUSTACEA; DIPLOPODA; INSECTA; PAUROPODA; SYMPHYLA. [C.B.C.]

Manganese

Chemical element number 25, *manganese*, Mn, has atomic weight 54.94. Manganese is one of the transition metals of the first long *period of the periodic* table and lies between chromium and iron. It has certain properties in common with both. Although relatively little known or used in its pure form, it has great practical importance in the manufacture of steel; about 14 lb of manganese is used for each ton of steel produced.



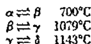
Ores and occurrence. The major ores of manganese are the oxides in hydrated and dehydrated forms, and to a lesser extent, the silicates and carbonates. Ores suitable for metallurgical purposes contain more than 40% manganese and must have a manganese-to-iron ratio of at least 9:1. World production of manganese ore is around 10,000,000 metric tons per year, and the principal producing countries in order of decreasing output are the Soviet Union, India, the Union of South Africa, Ghana, and French Morocco. The total production of these countries is about 8,000,000 tons of ore. The typical percentage analysis of a high-grade ore of Russian origin is MnO, 18.8; MnO_2, 62.4; Fe_2O_3, 1.3; SiO_2, 6.9; P_2O_5, 0.33; Al_2O_3, 1.1; CaO, 1.4; and H_2O, 5.5. This is equivalent to 53.97% Mn [illegible]

ess [illegible] of wrought steel and there is no effective substitute for it. It assists in the deoxidation of the steel bath and also, by combining with sulfur, markedly improves the hot-working properties of the steel. For this purpose, manganese is used in the form of ferromanganese, which contains approximately 80% manganese. It is used in high- and low-carbon grades, the former containing 6–7% carbon whereas the latter has only 0.1% carbon or less. High-carbon ferromanganese is made by the direct reduction of manganese ore with coke in a blast furnace. Ores of low silicon content containing at least 40% manganese and with a manganese-to-iron ratio of 9:10 are required for this purpose. The required ratio of manganese to iron is higher than might be thought necessary to produce an 80% manganese alloy because manganese has a high vapor pressure and consequently only 80% of it is recovered from the blast furnace, whereas the recovery of iron is practically 100%. Low-carbon ferromanganese is normally produced by the reduction of manganese ore with silicomanganese in an electric-arc furnace. The silicomanganese, which contains 69–71% manganese, 18–22% silicon, 4–6% iron, and less than 1% carbon, is also made in an

arc furnace by the treatment of ores of medium-iron and low-phosphorus content with carbon and silicon. Low-carbon ferromanganese naturally costs much more to produce than the high-carbon alloy but its use is essential to produce steels in which the carbon must be kept at a low level, for example, certain of the austenitic and heat-resistant steels. *See* FERROALLOY.

Production of pure manganese. Manganese of 97–98% purity is readily made on a commercial scale by the reduction with aluminum of high-grade manganese ore having a very low iron content. Manganese of even higher purity is made by the electrolysis of manganese(II) sulfate solution buffered with ammonium sulfate in a diaphragm cell employing lead-alloy anodes and alloy-steel cathodes. The manganese(II) sulfate is produced by leaching roasted high-purity ore with used catholyte solution from the electrolytic cell. This process, which is now operated on a commercial scale, produces manganese with less than 0.1% of metallic impurities at a cost comparable with that of low-carbon ferromanganese. Prior to the introduction of this process, high-purity manganese was a metallurgical rarity produced from the aluminothermic manganese of commercial purity by distillation and condensation in a vacuum.

Properties of pure manganese. Manganese melts at 1244 ± 3°C and boils at 2095°C. In the solid state it exists in four allotropic modifications, with the following transformation temperatures:



$$\alpha \rightleftharpoons \beta \quad 700°C$$
$$\beta \rightleftharpoons \gamma \quad 1079°C$$
$$\gamma \rightleftharpoons \delta \quad 1143°C$$

The $\alpha \rightleftharpoons \beta$ transformation is extremely sluggish and for this reason the transformation on heating appears to occur at appreciably higher temperatures than during cooling. The β-modification can be retained at room temperature by quenching from temperatures in the β-range. The γ-modification cannot, however, be completely retained on quenching. Electrodeposited manganese commonly assumes the γ-form in the as-deposited condition but transforms to the α-modification on standing at room temperature. δ-Manganese is stable only at high temperatures. The physical properties of manganese are summarized in the table.

Manganese in either the α- or β-modifications at room temperature is both hard and brittle and cannot be plastically deformed to any significant extent. Ductility is, however, associated with the simpler structure of γ-manganese. Manganese can be electrodeposited in this form in plating baths free from reduced sulfur compounds, and although it rapidly transforms to the α-form at room temperature, it can be stored by refrigeration for long periods of time without transformation. The ductility of the electrodeposited γ-manganese is such that sheet steel plated in this manner can be fabricated like tin or zinc-plated steel without damage to the manganese coating. Subsequent transformation to α-manganese is associated with a very considerable hardening to a level of 600–900 Brinell.

Manganese readily oxidizes in air to form a brown oxide coating. It also readily oxidizes at elevated temperatures. In this respect its behavior is more akin to its neighbor in the periodic table of higher atomic number, iron, than its neighbor with lower atomic number, chromium. A layered oxide scale is formed on oxidation in air, the layer nearest to the metal being MnO, and the outer layer, Mn_3O_4. The thickness of the MnO layer progressively increases, whereas that of the Mn_3O_4 layer decreases on oxidation at temperatures above 800°C. Below this temperature a third oxide layer, Mn_2O_3, occurs, and below about 450°C a fourth outermost layer of MnO_2 is stable.

Alloys of manganese. These can be classed conveniently as either ferrous or nonferrous alloys.

Ferrous alloys. Reference has already been made to the use of manganese as a deoxidant and desulfurizing additive in steelmaking. All commercial grades of steel, therefore, contain manganese as a minor but essential alloying constituent whose presence influences the cleanliness of the steel and its ability to be hot-worked. Small quantities of manganese also confer on the steel an improved resistance to impact, especially at low temperatures. Plain carbon, unalloyed steels normally contain about 0.3–0.7% manganese, resulting from its use in normal steelmaking procedures. Manganese is also used in somewhat larger quantities for low-alloy steels for constructional purposes and in high-duty cast irons. Rail steels commonly contain 0.9–1.2% carbon and there are other steels contain-

Physical properties of manganese

	α-Manganese	β-Manganese	γ-Manganese
Specific heat at 25°C, cal/(g)(°C)	0.114	0.155	0.120
Specific gravity at 20°C	7.21	7.29	7.21
Linear coefficient of expansion at 20°C	22.3×10^{-6}	24.9×10^{-6}	14.8×10^{-6}
Electrical resistivity, ohm-cm	$150\text{–}260 \times 10^{-6}$	90.0×10^{-6}	40.0×10^{-6}
Temperature coefficient of resistivity/°C	$2\text{–}3 \times 10^{-6}$	12.0×10^{-6}	60.0×10^{-6}
Crystal structure	Cubic	Cubic	Face-centered cubic; on cooling, it forms face-centered tetragonal crystals

ing 1.3–1.6% manganese. These steels have an increased tensile strength compared with steels of comparable carbon content and normal manganese contents and are produced in considerable tonnages.

The addition of manganese to iron lowers the temperature of the $\alpha \rightleftharpoons \gamma$ transformation of iron and with more than about 12% manganese the transformation is suppressed to below room temperature. These austenitic, nonmagnetic manganese steels have extremely important properties. As normally manufactured they contain 11–14% manganese and 1.0–1.4% carbon. They are used in the austenitic condition resulting from quenching from about 1050°C, at which temperature the carbides go into solution. In this condition they have a hardness of 230 Brinell and very unusual mechanical properties. With an ultimate tensile strength of 65–70 tons/in.², the elongation and reduction of area are 60 and 45% respectively. Unlike most strong metals, they show very little necking in a tensile test, the reduction of area being much the same all along the gage portion of the test piece. This is due to the fact that these steels show a very marked degree of work-hardening and this is the basis of their very successful use for parts, such as dredger buckets, track links, and rail crossings, which must withstand a high degree of shock and abrasion.

Manganese is also used in certain of the austenitic nickel-chromium steels, which have a wide use for corrosion- and heat-resistant applications. Steels of the 18% chromium, 8% nickel; and 25% chromium, 20% nickel types are very well known. Manganese can be substituted for some of the nickel in steels of this type without loss of their austenitic structure and their corrosion- and heat-resistant properties. Compositions of this type are Cr, 17%; Mn, 6%; Ni, 4%; Cr, 18%; Mn, 10%; Ni, 4%; and Cr, 16%, Ni, 15%; Mn, 6%; and Mo, 6%.

Nonferrous alloys. Manganese has a limited but nevertheless important use in nonferrous alloys. It is an effective deoxidizer of copper-base alloys and improves their mechanical properties. When added to brass containing 40–45% zinc, it increases the tensile strength at the rate of 0.7 tons/in.² per 1% of manganese, and 1% of manganese increases the elongation by 5%. Further additions decrease the ductility but continue to increase the strength. A brass containing 39% zinc and 1% manganese with small quantities of about 0.25% of iron and [illegible] used for marine propellers and low-temperature steam turbine blades.

Copper-manganese and copper-manganese-nickel alloys have very interesting properties. An alloy of the latter system containing approximately 84% copper, 4% nickel, and 12% manganese is a well-known alloy for electrical purposes (manganin). It has a resistivity of about 75×10^{-6} ohm-cm and a temperature coefficient of only 1×10^{-5} at 20°C. Other alloys in the system have even more remarkable properties; for example, an alloy of 60% manganese and 20% each of copper and nickel has a resistivity of about 190×10^{-6} ohm-cm combined with a temperature coefficient of resistance at least as low as that of manganin.

This alloy system also contains alloys with very high coefficients of thermal expansion. An alloy containing 72% manganese, 10% nickel, and 18% copper has a maximum expansion coefficient of 27×10^{-6}/°C and is made commercially as a high-expansion element of bimetals for thermostats.

Both aluminum and magnesium alloys commonly contain small quantities of manganese. Binary aluminum-manganese alloys containing 1.25–2.0% manganese have limited applications but manganese is also present to the extent of 0.5–0.6% in the age-hardening duralumin type of alloys and other precipitation hardening alloys where it has important effects in modifying the aging behavior and the corrosion resistance. *See* ALLOY; HEAT-TREATMENT (METALS AND ALLOYS); IRON ALLOYS; MANGANESE COMPOUND; METAL, MECHANICAL PROPERTIES OF; STEEL. [A.H.SU.]

Bibliography: C. A. Hampel (ed.), *Rare Metals Handbook*, 1954; A. H. Sully, *Manganese*, London, 1955.

Manganese compound

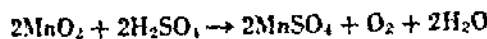
A combination of manganese with one or more other elements. Manganese is a fairly reactive metal. Although the massive metal is somewhat slow to react, the powdered metal reacts with ease and, in some cases, quite vigorously. When it is heated in air or oxygen, powdered manganese forms a red oxide, Mn_3O_4. With water, at room temperature, hydrogen and manganese(II) hydroxide, $Mn(OH)_2$, are formed. In the case of acids, because manganese is such a reactive metal, hydrogen is liberated along with the formation of a manganese(II) salt. Manganese reacts at elevated temperatures with the halogens, sulfur, nitrogen, carbon, silicon, phosphorus, and boron.

In its many different compounds, manganese has oxidation states of 1+ through 7+. The most common oxidation states are 2+, 4+, and 7+. All the compounds, except those containing Mn^{II}, are deeply colored. For example, potassium permanganate, $KMnO_4$, forms aqueous solutions that are a reddish-purple color; potassium manganate, K_2MnO_4, forms deep greenish-colored solutions.

Oxides. Manganese forms six oxides: MnO, Mn_3O_4, Mn_2O_3, MnO_2, MnO_3, and Mn_2O_7. The most common and best known are MnO, Mn_3O_4, MnO_2, and Mn_2O_7. The oxides that occur naturally are pyrolusite, MnO_2, braunite, Mn_2O_3, and hausmannite, Mn_3O_4. All the oxides are difficult to reduce to the free metal with carbon monoxide. They can be reduced only under pressure with hydrogen. However, lower oxides can be obtained, in some cases, by controlled thermal decomposition of more oxygen-rich oxides.

Manganese dioxide, MnO_2, which has a gray to gray-black color, is by far the most important oxide. Although it is usually found in the free state, manganese dioxide can be prepared by heating manganese nitrate, $Mn(NO_3)_2$, at 180–200°C, or by heating a mixture of manganese carbonate, $MnCO_3$, and potassium chlorate at 300°C. There is some question about the purity of the product obtained; it has been stated that the purest product contained only 98% MnO_2. The discrepancy was attributed to the formation of a mixture of MnO_2 and MnO.

The oxide is insoluble in water and in weak acids and bases. However, it does react with concentrated sulfuric acid:



$$2MnO_2 + 2H_2SO_4 \rightarrow 2MnSO_4 + O_2 + 2H_2O$$

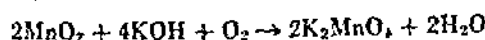
and with a reducing acid, such as concentrated hydrochloric acid, at room temperature or slightly higher:



$$MnO_2 + 4HCl \rightarrow MnCl_2 + Cl_2 + 2H_2O$$

With ice-cold concentrated hydrochloric acid, it gives a red- to brown-colored solution which probably contains manganese(III) chloride, $MnCl_3$, and manganese(IV) chloride, $MnCl_4$.

In the presence of air or other oxidizing agents, MnO_2 reacts with fused potassium hydroxide to form potassium manganate, K_2MnO_4:



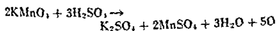
$$2MnO_2 + 4KOH + O_2 \rightarrow 2K_2MnO_4 + 2H_2O$$

In weakly acid solution, potassium manganate disproportionates to form, among other products, potassium permanganate, $KMnO_4$.



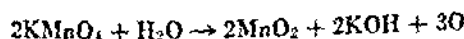
$$3K_2MnO_4 + H_2O \rightarrow 2KMnO_4 + MnO_2 + 4KOH$$

Potassium permanganate is a powerful oxidizing agent, both in organic and inorganic chemistry. In acid solution, five oxygen atoms are liberated by each two formula weights that react:



$$2KMnO_4 + 3H_2SO_4 \rightarrow K_2SO_4 + 2MnSO_4 + 3H_2O + 5O$$

whereas in basic solution, only three oxygen atoms are liberated:



$$2KMnO_4 + H_2O \rightarrow 2MnO_2 + 2KOH + 3O$$

The compound is also useful as an oxidizing agent in the analytical chemistry of certain metal ions.

Halides. The most common manganese compounds with fluorine, chlorine, bromine, and iodine are those containing manganese in a 2+ oxidation state. These compounds, having the formulas, MnF_2, $MnCl_2$, $MnBr_2$, and MnI_2, respectively, can generally be prepared by the direct combination of manganese metal with the halogen.

The most useful of the halides are manganese(II) fluoride, MnF_2, and manganese(II) chloride, $MnCl_2$. The fluoride is best prepared by heating NH_4MnF_3 in an atmosphere of carbon dioxide. It is a colorless compound which is slightly soluble in water. The pink-colored chloride can be prepared by the reaction of manganese [illegible] warm concentrated hydrochloric acid [illegible] tion of dry hydrogen chloride on MnO, [illegible] manganese metal.

Other compounds. In the presence of [illegible] bases, solutions of manganese(II) [illegible] pink-colored, insoluble manganese(II) [illegible] In the presence of [illegible]

$MnO_2 \cdot nH_2O$ [illegible] its insolubility in water, is a [illegible] When heated in the absence of air, [illegible] whereas in air, the main product [illegible]

Ammonium sulfide or alkali [illegible] cipitate pink-colored manganese(II) [illegible] solutions of manganese(II) [illegible] tated from solution on [illegible] potassium hydrogen carbonate. It [illegible] posed on heating or by treatment [illegible]

Other important manganese(II) [illegible] pink-colored manganese(II) [illegible] pink-colored manganese(II) [illegible] colorless manganese(II) [illegible]

Uses. The compounds of [illegible] uses in industry. Manganese [illegible] drying agent or catalyst in [illegible] a decolorizer in glass [illegible] cells. Potassium permanganate [illegible] ing purposes, decolorization [illegible] dizing agent in preparative [illegible] istry *See* DRY CELL; [illegible]

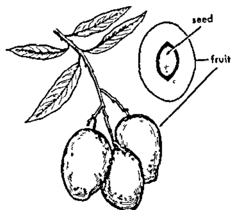
Analytical methods. [illegible] determined gravimetrically [illegible] manganese ammonium [illegible] H_2O, followed by ignition [illegible] phate, $Mn_2P_2O_7$, which [illegible] manganese from certain [illegible] fide forms insoluble [illegible] qualitative test for [illegible] quantity of the compound [illegible] Green-colored manganate [illegible] in the reaction and [illegible] MANGANESE; TRANSITION [illegible]

Manganite

A mineral having [illegible] crystallizing in the [illegible] matic crystals with [illegible] ganite crystals are [illegible] masses. There is [illegible] ness is 4 (Mohs [illegible] 4.3. The luster is [illegible] Manganite is [illegible] oxides, and pseudomorphs [illegible] ganite are common. [illegible] in the Harz Mountains, [illegible] the United States [illegible] nite is a minor [illegible] PYROLUSITE.

Mango

This plant, *Mangifera indica*, is a member of the sumac family (Anacardiaceae) and a native of southern Asia, but it is now grown in Indomalaysia, Africa, tropical America, California, and Florida. The mango has been in cultivation for almost 6000



Mango (*Mangifera indica*) developed in Florida and California. (From C. J Hylander and O. B Stanley, *College Botany*, Macmillan, 1949)

years. It is an important fruit in tropical countries throughout the world, more important than peaches and apples among temperate climate fruits. It is eaten by at least one-fifth of the world's inhabitants. The fruit is a fleshy drupe, 3–5 in. long, having a thick, yellowish-red skin and a large seed. The mango is one of the few tropical fruits which have been improved under cultivation, and over 1000 different varieties are now being grown. *See* SAPINDALES. [P.D.S.]

Mangrove swamp

A swamp forest of low to tall trees and some shrubs, commonly associated with some salt marsh herbs. Swamp forests occur along the borders of many tropical shores where wave action is not intense and mud and peat are deposited (Fig. 1*a*). Most plants of this community are halophytes that are well adapted to salt water and fluctuations of tide level. Some have stilt or prop roots to help hold them on the shifting sediments and others have erect root structures (pneumatophores) that crop out above the surface (Fig. 1*b*, *c*). Several species have well-developed vivipary of their seeds, the hypocotyl developing while the fruit is held on the tree. These seedlings are usually so shaped and weighted that they float long distances in the sea and thus extensive migration is ensured.

The mangrove community often develops as a distinct halosere. In such areas it is zoned from open water landward in a series of different species (Fig. 2). The landward zone species develop on sediments and peats which were initially deposited in the seaward zone. These changes due to deposition in the swamp often extend the coast outward and form incipient islands in shallow, quiet waters. The swamps when dense also afford some protection against erosion resulting from violent storms. Thus, mangrove swamps have a significant geologic role. *See* CLIMAX PLANT FORMATIONS; ECOLOGY.

The composition of mangrove swamps is strikingly different in two regions. Only four species predominate in the Americas and along the west coast of Africa, but over 10 species are frequent in swamps of eastern Africa, Asia, and the western Pacific region. Important genera are *Rhizophora*,







Fig. 1. Mangrove swamps, Great Barrier Reef. (a) Detail of mangroves on north side of Howick Island. (b) Interior of mangrove swamp, Newton Island. (c) Mangroves at northwestern corner of King Island. (From J. A. Steers, Salt marshes, *Endeavour*, 18(70):75–82, 1959)

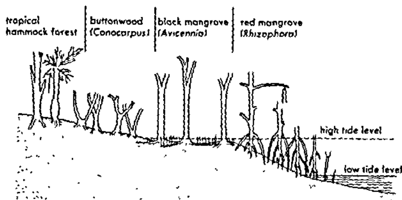


Fig. 2. Zonation of mangroves and associated vegetation, in Florida. (After J. H. Davis, The Ecology and Geologic Role of Mangroves in Florida, Carnegie Inst. Wash. Publ. 517, 1940)

Avicennia, Laguncularia, Bruguiera, Sonneratia, Carapa, and *Ceriops*. These species belong to a number of families. They are examples of convergent adaptations to their halic habitat.

Mangroves are used, especially in the Orient, for some construction, charcoal, and tannin. The forest is one of the few types that will replace the same species after exploitation. [J.H.D.]

Manic depressive psychosis

A severe disturbance of affect (feeling or emotion); hence it is referred to as an affective psychosis. It is characterized by extreme and pathological elation alternating with severe dejection, both of which may last for months or years. There may be normal intervals between these states, and cases are known in which only the manic or depressive attack occurs once or repeatedly. *See* PSYCHOSIS.

The condition is relatively common. No exact figures exist, but mental hospitals report as manic depressive psychosis about 5–15% of their admissions. Frequently the diagnoses of patients with manic depressive psychoses are changed to schizophrenia after there have been several attacks of disturbed behavior. *See* SCHIZOPHRENIA.

During the manic phase, the patients are euphoric (have an exaggerated sense of well-being) and overactive. They talk incessantly and show a flight of ideas, but their thoughts are not illogical or bizarre. In extreme cases patients become agitated, and many exhaust themselves, especially when they do not sleep or eat sufficiently. Manic patients are also prone to be a nuisance to their fellow human beings and do occasionally commit sexual and aggressive crimes.

The depressive phase is characterized by sadness, grief, anxiety, and low self-esteem. Patients feel guilty over minor or imaginary unsocial behavior. They are also apt to have very severe hypochondriacal delusions. Their thought processes are retarded, but not disordered. Depressive patients eat and sleep poorly and usually lose weight. The most serious symptom is the inclination of such patients to commit suicide, particularly when the patient is about to come out of the depth of his depression, because at such a time the psychic (mental) retardation is less pronounced.

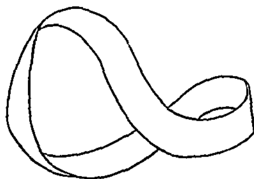
The etiology of the disorder is obscure. Endocrine and constitutional factors probably play a role, although specific data to prove this are lacking. Hereditary factors, as shown in research on twins, are of importance. The meaning of symptoms has been elucidated by psychoanalytic investigators who follow Freudian lines of thought. In manic states the ego and superego are overwhelmed by the id. The id is the reservoir of instincts or impulses; the ego is that part of the personality dealing with adaptations to the external world; the superego is commonly called conscience and restrains the id and ego. In depressed states, which are qualitatively similar to normal mourning, the id and ego are enslaved by a tyrannical superego. The differential diagnosis must consider organic reactions, schizophrenia, and involutional reactions. There are also borderline conditions referred to as hypomanic states. *See* PSYCHOANALYSIS.

In depressive states, electric convulsive therapy in connection with psychotherapy has been quite successful. Drug treatment with certain drugs, known as psychic energizers, has shown promise. Manic states are best treated by tranquilizing drugs and psychotherapy. Hospitalization is usually necessary. *See* PSYCHIC ENERGIZER; TRANQUILIZER. [F.C.R.]

Bibliography: J. R. Ewalt, E. A. Strecker, and F. G. Ebaugh, *Practical Clinical Psychiatry*, 8th ed., 1957.

Manifold

An *n*-dimensional manifold is a connected, locally compact space with a countable basis, each point of which has a neighborhood homeomorphic to euclidean *n*-space (see TOPOLOGY). Examples are simple closed curves (1-dimensional manifold) and surface of a sphere or a torus (2-dimensional manifolds). A fundamental problem of topology (solved only for $n = 1,2$) is to classify *n*-dimensional manifolds into types such that two manifolds are homeomorphic if and only if they belong to the same type. If a long rectangular ribbon be given half a



Möbius band.

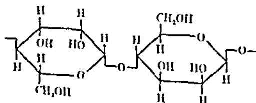
turn and the two ends glued together, the resulting surface is called a Mobius band. A manifold is orientable provided no portion of it is homeomorphic to a Mobius band; otherwise, it is nonorientable. A manifold has connectivity k if the maximum number of simple closed curves that may be traced upon it without decomposing it into two or more pieces is $k - 1$. For examples a sphere has connectivity 1, and a torus, connectivity 3. The type of a 2-manifold is determined by its connectivity and orientability. If ρ holes are drilled in a sphere, the resulting surface has connectivity $k = 2\rho + 1$. The number ρ is the genus of the surface. The most general 2-dimensional orientable manifold is obtained in this way. [L.M.BL.]

Manifold, engine

In mechanics, a pipe fitting that provides multiple connections to or from a single pipe. When applied to a multicylinder engine, one manifold distributes air or an air and fuel mixture from the intake to the cylinders (intake manifold), and another manifold collects exhaust gas from the cylinders and delivers it to a muffler and exhaust pipe (exhaust manifold). A particular requirement of the intake manifold is that it distribute equal quantities of air and fuel to all cylinders. Both the intake and exhaust manifolds should preferably present low impedances to the gas flow into and out of the cylinders. See INTERNAL COMBUSTION ENGINE. [N.M.]

Mannan

One of a group of polysaccharides composed chiefly or entirely of D-mannose units. Mannans occur in the thickened cell walls of palm seeds as reserve polysaccharides, which disappear during germination. Mannans are generally prominent components of coniferous woods, but they occur in only very small quantities in deciduous woods. Ivory-nut mannan is the alkali-soluble polysaccharide composed of 1 → 4-linked β-D-mannose units. The molecule is linear and is somewhat similar to cellulose.



Salep mannan, extracted from the tubers of various orchids, is similar in structure to ivory-nut mannan. This mannan appears to be readily digested. Yeast mannan, commonly called yeast gum, is highly branched. Both this polysaccharide and glycogen are prominent components of the yeast cell. In yeast mannan, one- or two-membered side chains, linked 1 → 6, extend from a long main chain. Mannocarolose is the mannan formed when *Penicillium charlesii* G. Sm. is grown on Czapek-Dox solution of D-glucose. The mannan of *Porphyra umbilicalis*, a red alga, contains 1 branch unit for each 12 D-mannose units.

Galactomannans, composed of D-galactose and D-mannose units, are water-soluble and form highly viscous solutions. These plant mucilages are reserve carbohydrates in the endosperm of leguminous seeds. Guaran, a product of guar seed endosperms, is composed of a main chain of D-mannose units linked 1 → 4 with single D-galactose units linked 1 → 6 as side chains on every other D-mannose unit. Guaran is used industrially in paper, mining, textile, leather, and food manufacture. Locust bean gum differs from guaran only in the number of D-galactose units attached as side chains. This galactomannan is obtained from the locust (carob) bean, which is grown in the Mediterranean region and is used as a food called St. John's bread. Locust bean gum is widely used for the sizing of textiles and paper, for the preparation of printing and dyeing mixtures, and as a thickener in the preparation of foodstuffs, pharmaceutical jellies, and cosmetics. Alfalfa seed gum, the endosperm of the alfalfa seed, is composed of a chain of alternating D-galactose and D-mannose units to which single D-galactose units are linked as side chains. The ratio of D-galactose to D-mannose units in this galactomannan is 2:1.

Glucomannans are the principal polysaccharides in the corms of various species of *Amorphophallus*. Konjak mannan is the glucomannan found as the chief component of flour prepared in the Far East from tuberous corms of *Amorphophallus konjak* (*rivieri*). The corms of *Amorphophallus oncophyllus* yield a glucomannan which is an excellent beater additive in paper manufacture. Other glucomannans occur in the cell walls of certain trees. See CELL WALLS IN PLANTS; GUM; MONOSACCHARIDE; POLYSACCHARIDE. [A.M.B.; R.L.WH.]

Manometer

A double-leg liquid column gage used to measure the difference between two fluid pressures. The manometer is the basic standard for measurement of small differential pressures.

Two principal varieties are the glass-tube manometer, for simple indication of difference of pressure, and the metallic-housed mercury manometer, used to record or control difference of pressure or fluid flow. The three basic glass-tube types

the U-tube, well-type, and inclined-tube manometers, may be used to measure vacuum or gage pressure by leaving *one* side open to atmosphere.

U-tube manometer. If the legs of the manometer are connected to separate sources of pressure, the liquid will rise in the leg with the lower pressure and drop in the other leg. The difference between the levels is a function of the applied pressures and the specific gravity of the instrument fluid. The cross-sectional area of the tubes does not affect the difference between the levels. A scale graduated in inches or centimeters is commonly affixed between the legs of the manometer.

Micromanometer U tubes have been made using precision-bore glass tubing, a metallic float in one leg, and an inductive coil to sense the position of the float. A null-balance electronic indicator can detect pressure changes as minute as 0.0005 in. of water. Such devices are used only as laboratory standards.

Well-type manometer. One leg of this type has a relatively small diameter; the second leg is a reservoir. The cross-sectional area of the reservoir may be as much as 1500 times that of the vertical leg, so that the level of the reservoir does not change appreciably with change of pressure. Small adjustments to the scale of the vertical leg compensate for the little reservoir level change that does occur. Readings of differential or gage pressure may then be made directly on the vertical scale. Mercurial barometers are commonly made as well-type manometers.

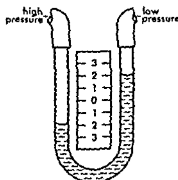


Fig. 1. U-tube manometer.

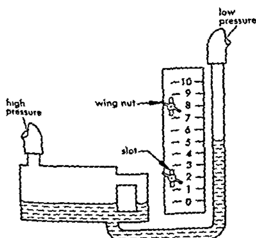


Fig. 2. Well-type manometer with zeroing adjustment.

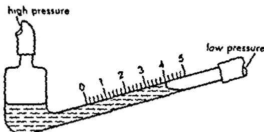


Fig. 3. Inclined-tube manometer.

Inclined-tube manometers. These are used for gage pressures below 10 in. of water column. The leg of the well-type manometer is inclined from the vertical to elongate the scale. Inclined double leg U-tube manometers are also used to measure very low differential pressures.

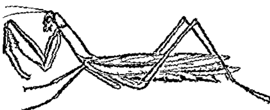
Glass-tube manometers, while considered accurate and repeatable, do not produce mechanical motion to actuate recording and controlling mechanisms. For such applications, mercury manometer float-type meters, bell-type gages, or diaphragm gages are employed. *See* FLOW MEASUREMENT; PRESSURE MEASUREMENT.

The U-tube and the well-type manometers have an accuracy of about 0.05 in. The inclined-tube manometer, with its longer column, is accurate to about 0.01 in. The accuracy depends on the operator's skill and the cleanness of the liquid and tube. [B.D.H.; H.G.P.]

Mantis

Any member of the insect family Mantidae, order Orthoptera. Mantids, often called praying mantis, are usually considered the only beneficial insects of the order Orthoptera because they are predatory and feed upon other insects. They are primarily tropical animals, but a few range well into the temperate regions. Of the 600 known species about 25 occur in North America.

Mantids are well known for their predatory habits, and for their tendency toward cannibalism. They are also well known for their concealing coloration, being green, gray, brown, or a mixture



The praying mantis, *Mantis religiosa*; length to 2 in. (From *E. L. Palmer, Fieldbook of Natural History*, McGraw-Hill, 1949)

of green and gray, and so marked as to be almost invisible to unsuspecting insects. The body shape of many imitates the forms of the leaves of the plants in which they hide.

The name praying mantis is derived from the prayerful posture they assume with their powerful front legs raised, waiting to strike a victim.

The mantids overwinter in the egg stage; the eggs are laid in a strong, grayish ootheca, glued to any upright support. There is only one generation each year. *See* ORTHOPTERA. [J.D.B.]

Manure

Any plant and animal residue that may contain excreta of animals. Manures, classified into animal manure, plant manure, and compost, are added to the soil in various stages of decomposition. In the soil they undergo further degradation through the action of soil microorganisms, raising the soil fertility level, and improving soil texture by increasing humus content (*see* HUMUS).

Animal manure includes plant residues like straw, used as litter, as well as solid and liquid excreta. Microorganisms attack the nitrogenous components and the more readily fermentable carbohydrates during storage of the manure. Intense microbial respiration, favored by a loose texture and adequate moisture in the manure, causes a rise in temperature so that thermophilic forms of bacteria, actinomycetes, and fungi grow.

Compost consists of plant and animal residues allowed to rot before being applied to soil. Because they have a lower ratio of nitrogenous to carbohydrate material than animal manure, composts are usually reinforced with inorganic nitrogen and available phosphorus to facilitate microbial action. With moisture and aeration, decomposition proceeds accompanied by a rise in temperature.

Green manure is plant material in the form of a growing crop plowed into the soil. Because the object of manuring is to increase the supply of nitrogen, leguminous crops are grown and then plowed into the soil. Green plants are higher in soluble carbohydrates, nitrogen, and minerals than plant residues used in manures and composts; as a result, decomposition sets in more rapidly. Green manures, being low in cellulose and lignin, have little effect on the humus content of soil. *See* BACTERIAL METABOLISM; SOIL MICROBIOLOGY. [A.C.L.]

Map design

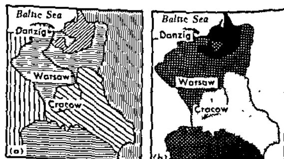
The manner of delineation and arrangement of all the visual components of a map, consisting of lines, [illegible]

[illegible] aesthetic matter, the design of maps has become of great significance as a result of increasing recognition requirements such as in military maps and aeronautical charts. The development of design standards has not kept pace with user requirements.

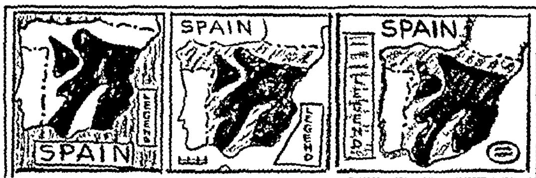
The role of convention is strong in map design partly because familiarity tends to establish the status quo, but also because most of the techniques and media used in cartography, such as color, became well established before the tremendous upsurge in cartographic activity in the twentieth century. Consequently, many cartographic techniques are so firmly entrenched that rapid replacement is unlikely. Nevertheless, a marked change in attitude toward map design and tradition has come about during the first half of the twentieth century. Much of this is due to the increasing freedom occasioned by developments in the map reproduction processes and to the development of psychometric testing methods. *See* MAP REPRODUCTION.

Prior to the first part of the twentieth century most maps were reproduced by engraving methods, and several hundred years' familiarity with the kinds of lines, shapes, and patterns easily produced by the burin and etching techniques developed a strong tradition. The development of lithography and the modern photolithographic processes opened the door to many other possibilities. Similarly the development of psychometric testing techniques provided the cartographer a means whereby he could evaluate the relative effectiveness of the symbolization used on maps. New kinds of maps to be used for new purposes made necessary new kinds of design, as for example the requirement of rapid recognition in aeronautical charts used for high-speed flight. Another factor of importance has been the rapid increase in the variety of media available to the cartographer. Most of the changes in design have been in the fields of map symbolization, lettering, and color.

Symbolization of the various classes of map data has undergone close scrutiny. The rise of competitive road maps since the inception of large-scale automobile travel has resulted in a steady development of their design. The availability of preprinted symbols for road number designations and for all other kinds of map symbols has led to more uniformity and a considerably higher average design quality. Tests of symbol effectiveness by scholars and government agencies have become available in the literature, leading toward higher standards gen-



Sketch maps employing: (a) line patterns, (b) dot patterns. (From A. H. Robinson, *Elements of Cartography*, 2d ed., Wiley, 1960)



Thumbnail sketches of a map made in the preliminary stages to arrive at a proper balance. (From A. H. Robinson, *Elements of Cartography*, 2d ed., Wiley, 1960)

erally. The most generally significant change in symbolization has been the increasingly graphic representation of terrain on all maps and especially its visual representation on standard topographic series in combination with contour lines. The increasing use of maps presenting statistics has resulted in the testing and devising of many complex symbol forms with which to show several classes of data on the same map.

The design of lettering for maps has consistently changed for the better since the beginning of the twentieth century. Formerly the engraver was responsible for most lettering, but photoprocessing of drafted copy enabled the cartographer to choose any sort of lettering. The development of preprinted lettering opened the typographer's entire stock to the cartographer. After the first rush of overlettering due to availability, the tendency was to return to the simplicity characteristic of hand-lettered copy. No lettering styles designed especially for cartographic use have yet been generally accepted.

The greatest changes have occurred in color use as a result of the advances in color science and the increasing versatility of the reproduction processes. Halftones and color printing being now widely available, more subtle color combinations and pastels are coming into favor. A greater proportion of maps is being printed in color than ever before.

In the history of cartography no greater changes have taken place than those now underway in the field of map design. Formerly it was primarily aesthetic; it is now assuming the status of a science. See CARTOGRAPHY. [A.H.R.]

Bibliography: A. H. Robinson, *The Look of Maps*, 1952.

Map projections

Methods of transforming a spherical representation of the earth's surface to a nonspherical surface, usually a plane. These methods are useful because globes and curved-surface reproductions are cumbersome, expensive, and difficult to measure upon. An infinite number of methods is possible. Each method deforms the geometric relationships among points on the sphere in some way, since a sphere and a plane are not applicable surfaces. Consequently, directions, distances, areas, and angular relationships on the earth are never completely recreated on a map.

Deformation patterns. Transformation of the spherical surface may be accomplished in two ways: (1) by geometric transfer to some other surface, such as a tangent or intersecting cylinder, cone, or plane, which can then be developed, that is, cut apart and laid out flat; or (2) by direct mathematical transfer to a plane of the directions and distances among points on the sphere. Whether a projection is geometrically or mathematically derived, if its pattern of scale variation is like that which results from geometric transfer, it is classed as cylindrical, conic or, in the case of a plane, azimuthal or zenithal. See CARTOGRAPHY; COORDINATE SYSTEMS, TERRESTRIAL; MAP SCALE.

Cylindrical projections. These result from symmetrical transfer of the spherical surface to a tangent or intersecting cylinder. True or correct scale will obtain along the great circle of tangency or the two homothetic small circles of intersection. If the axis of the cylinder is made parallel to the axis of the earth the parallels and meridians will appear as right, perpendicular lines. Points on the earth equally distant from the tangent great circle (Equator) or small circles of intersection (parallels equally spaced on either side of the Equator) will have equal scale departure. The pattern of deformation will therefore parallel the parallels, as change in scale occurs in a direction perpendicular to the parallels. If the cylinder is turned 90° with respect to the earth's axis the projection is said to be transverse, and the pattern of deformation will be symmetric with respect to a great circle through the poles. If the turn of the cylinder is less than 90°, an *oblique* projection results. All cylindrical projections, whether geometrically or mathematically derived, have similar patterns of deformation.

Conic projections. Transfer to a tangent or intersecting cone is the basis of these projections. True scale will obtain along one or two small circles in the same hemisphere. Conic projections are usua

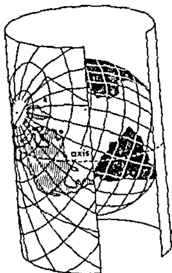


Fig. 1. Oblique Mercator projection. (*From American Oxford Atlas, Oxford, 1951*)

arranged with the axis of the cone parallel to the earth's axis. Consequently, meridians appear as radiating straight lines and parallels as concentric arcs which intersect with the meridians at right angles. Conical patterns of deformation parallel the parallels, that is, scale departure will be uniform along any parallel. Several important conical projections are not true conics in that their derivation either is based upon more than one cone (polyconic) or is based upon one cone with a subsequent rearrangement of scale variation.

Azimuthal projections These result from the transfer to a tangent or intersecting plane established perpendicular to a right line passing through the center of the earth. All geometrically developed azimuthal projections are transferred from some point on this line. Points on the earth equidistant from the point of tangency or the center of the circle of intersection have equal scale departure. Hence the pattern of deformation is circular and concentric to the aforementioned center. All azimuthal projections, whether geometrically or mathematically derived, have two aspects in common: (1) all great circles that pass through the center of the projection appear as straight lines, and (2) all azimuths from the center are truly displayed.

Nongeometric projections show *no consistency* in their patterns of deformation, but many are useful, *nevertheless*. One such group, called oval projections, is derived by mathematically arranging the spherical surface within a shape bordered by a series of specified types of curves, such as ellipses, parabolas, or sine or quartic curves.

Properties of projections. The retention anywhere on the projection of some specif [illegible]

[illegible]

Many single properties or combinations of properties are possible. The use to which a map is to be put determines the properties to be retained or created in the transformation process. All properties are the result of arranging the magnitude and direction of scale variation. Properties most widely employed are conformality, equivalence, azimuthality, and the transformation of significant lines on the earth to simple lines on the map.

Conformality (also orthomorphism) is the retention of angular relationships at each point. It is obtained by arranging the scale so that at each point the scale is uniform in all directions. Since the spheroid and plane are not applicable, scale must change from point to point. Thus, all conformal projections deform the relative sizes of earth areas. Conformal projections are most widely used for navigational, engineering, and topographic maps since observable angles may be measured on the map with a protractor.

Equivalence (also equal-area) is the retention on a projection of the relative sizes of areas. It is obtained by arranging the scale at each point so that the product of the sides of an infinitesimal earth rectangle projected anywhere on the map is the same. Since the two surfaces are nonapplicable, the scale at a point must be different in different directions. Hence the properties of conformality and equivalence are mutually exclusive. Equivalent projections are most widely used for mapping geographic, economic, and similar types of data whenever the areal extent of the phenomena is important. The extent of earth area within administrative subdivisions can best be derived by measurement from equivalent projections.

Azimuthality is the retention of azimuths (correct directional relations along a great circle bearing) from one point to another. Within limited areas this, like all properties, can be approached on many projections, but can be precisely retained from only one or two points. The centers of all projections theoretically or actually constructed on a tangent plane have this property. Azimuthality is useful in graphic aids to communication and for observations dealing with electronic impulses that travel along great circles.

Significant lines on the earth include such useful concepts as great circles, loxodromes, and small circles, which are utilized in navigation, communication, and area and space analysis. By arranging the

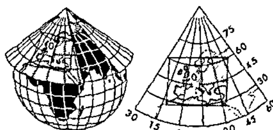


Fig. 2. Conical projection, based on origin 45°N, 10°E. (*From American Oxford Atlas, Oxford, 1951*)

scale, these may be made to appear as straight or consistently curved lines on a map. Of special significance are loxodromes and great circles in air and sea navigation. The economical direction to move across the earth is along a great circle. The practical requirement of guiding movement by compass angle requires that headings be along loxodromes (rhumb lines). Scale on projections can be arranged so that entire families of these kinds of critical lines may be made to appear as straight lines. Similarly, directional concepts such as a specific cardinal direction can be made parallel anywhere on the map. Any circle on the earth can be made to appear as a circle on the map. Many other similar attributes are possible with map projections.

Select examples. Although an infinite number of different projections is possible only a relatively few are widely used. The commonest projections and families of similar projections are the Mercator, gnomonic azimuthal, stereographic azimuthal, Lambert conformal conic, Albers equivalent conic, and polyconic.

The Mercator projection is a mathematically derived, cylindrical-type, conformal projection which is conventionally based on the Equator as the tangent great circle. The linear scale increases in a direction perpendicular to the Equator as the secant of the latitude on a sphere. All loxodromes are straight lines on the conventional form, making the Mercator the most widely used projection for nautical navigation. The transverse Mercator (and similar variants for the spheroid, for example, Gauss-Krüger) employs a meridian pair as the tangent great circle. The transverse forms are widely used as frameworks for topographic maps and military grid systems. Straight-line loxodromes do not obtain on the transverse forms.

The gnomonic azimuthal projection is obtained by geometric transfer from the center of the sphere to a tangent plane. All arcs of great circles appear as straight lines. It is widely used in navigation to derive courses which when plotted on the Mercator can be approximated by a series of headings along rhumb lines.

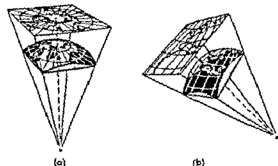


Fig. 3. Azimuthal projections. For clarity of presentation, tangent planes have been raised above sphere surface. (a) Polar azimuthal (b) Oblique azimuthal (From American Oxford Atlas, Oxford, 1951)

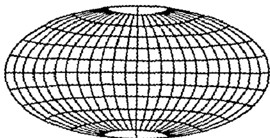


Fig. 4. Grid for Aitoff-Hammer oval world projection. (From American Oxford Atlas, Oxford, 1951)

The stereographic azimuthal conformal projection is derived by geometric transfer to a tangent plane from a point opposite the point of tangency. All circles on the earth appear as circles on the map. It is widely used as a framework for topographic maps of small areas and for navigation and grid systems in polar areas.

The Lambert conformal conic with two true scale (standard) parallels is widely used for aeronautical and meteorological charts, as a framework for topographic series of smaller countries, and for rectangular grid systems of smaller areas. Scale increases away from the standard parallels, and great circles are nearly straight lines.

The Albers equivalent conic with two standard parallels is similar in appearance to the Lambert. It is used in middle latitudes for administrative and area research maps.

The polyconic (the ordinary, or American, polyconic) has no specific property. Each parallel is true scale It is used for large-scale topographic maps, each sheet being individually developed. Sheets fit together north-south but not east-west. The modified polyconic employs straight meridians and distributes scale error over the map sheet. It is selected as framework for the International Map of the World.

Most other forms of projection are used for small-scale, general reference, special-purpose maps. Most widely used for these are equivalent or near-equivalent varieties. [A.H.R.]

Bibliography: C. H. Deetz, O. S. Adams, *Elements of Map Projection*, USCGS Spec. Publ. 68, 1921; A. H. Robinson, *Elements of Cartography*, 2d ed., 1960; J. A. Steers, *An Introduction to the Study of Map Projections*, 10th ed., 1956.

Map reproduction

The processes of duplicating map drawings. Reproduction is accomplished by many techniques, each appropriate in terms of cost and quality to the purpose of the map. Many of the processes may also be used in intermediate stages in the preparation of a map. The reproduction process to be employed has considerable effect upon the techniques and media used in the preparation of the map drawing.

The major distinction among processes is between those which use ink on a printing plate and

those which employ light-sensitive emulsions on paper. The former are appropriate for large runs since unit cost per copy decreases rapidly as quantity printed increases. The latter are usually preferred for short runs and intermediate stages in the cartographic drafting procedure.

Inked-plate printing processes. These include three major varieties: a relief printing plate (letterpress), an intaglio printing plate (engraving), and a planographic printing plate (lithography). Engraving, formerly the most commonly used reproduction technique for maps, has largely been discarded in favor of the others due to its cost and lack of versatility. Both letterpress and lithography involve similar photographic processing of the original drawing (called the copy). Copy to be reproduced is placed in a vacuum frame rigidly fixed to the camera The vacuum frame may be moved with respect to the lens in order to change the size of the image on the negative in the camera. Ordinarily copy is drawn at a larger size for reduction in the camera. The film in the camera is slow orthochromatic on a stable film base. It is sensitive to the red end of the spectrum but not to the blue; consequently instructions and guide lines on the copy may be rendered in blue and need not be removed prior to photography. The film is wet developed and no continuous tone appears; it consists only of opaque or translucent sections. Retouching is done to close pinholes, and modifications may be accomplished by stripping, opaquing, or scraping the emulsion.

The printing plate is prepared by placing the film in contact with the sensitized plate and exposing it to arc lights The light-sensitive emulsions vary depending upon whether the subsequent processing is to harden the surfaces exposed or those not exposed through the translucent sections of the negative. Processing of the printing plate after exposure consists of etching the nonprinting areas on the letterpress plate, or making ink-receptive the printing areas on a planographic plate. After processing, the plate is placed upon a press, usually rotary, and is inked before each impression. In order to conserve the life of the printing plate in lithography, the plate commonly transfers the image to a soft rubber roller which [illegible]

pa[illegible]

[illegible]

gray scale as in a panchromatic photograph) is obtained in either process by inserting a halftone screen immediately in front of the negative in the camera to break the light into uniformly spaced dots varying in size in proportion to the amount of light reflected from the different parts of the copy.

Color printing of maps is accomplished in two ways regardless of whether the printing process is letterpress or lithography, flat color and process color. In flat color a separation drawing (or negative) and plate are made for each color to be used on the press. In process color, copy consists of one black copy and one colored copy containing all the colors in combination, as in a painting. By means of filters on the lens the proportions of each of the primary colors (red, blue, yellow) in the mixtures on the copy are translated into halftone negatives which are then printed successively. The greatest problem in color printing is the maintenance of the positional accuracy of the successive impressions, or their register. The difficulty of obtaining register of separate drawings in the flat-color process has led to two major developments since 1940 —the use of plastics as drafting surfaces and the direct preparation of negatives for platemaking. Plastics, especially the vinyl series, are dimensionally stable whereas paper is not. The direct engraving (scribing) on film or glass in optically translucent but actinically opaque emulsions makes unnecessary the photographic stage in map reproduction.

Non-ink reproduction processes. A large variety of non-ink processes, such as photostat and various diazo processes, are used mostly in intermediate stages of cartography. Copy for photographic processing may be on opaque or translucent media and the reproductions may be enlarged or reduced They usually require a film- or paper-negative stage and wet developing. Diazo processes require translucent copy and cannot be enlarged or reduced They may be either wet or dry developed and may produce directly either negative or positive copies. *See* CARTOGRAPHY. [A.H.R.]

Bibliography: A. H. Robinson, *Elements of Cartography*, 2d ed., 1960.

Map scale

The relation between the length of a line on the map to the corresponding line in nature. Map scale is expressed in three ways:

1. Representative fraction, as

$$\frac{1}{R} = \frac{\text{length on map}}{\text{length in nature}}$$

Thus, if on a map, 1 inch = 10 miles, it is expressed by the representative fraction 1/633,600 or 1:633,600 as there are 63,360 inches in a mile.

2. Miles-per-inch scale indicates the number of miles represented and is shown on maps as

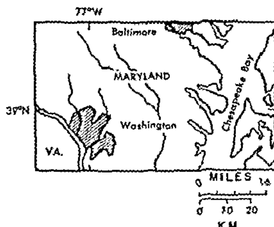
Scale: 1 inch to 10 miles

3. Graphic scale is a graduated line that portrays the actual length of miles or kilometers as they appear on the map. Graphic scales have the great advantage of remaining true after the map has been enlarged or reduced in reproduction. (See illustration.)

In countries using the metric system the common scales are 1:25,000, 1:50,000, 1:100,000, and so on. British maps generally use scales based on multiples of 1:63,360; for example, the so-called quarter-inch maps which mean

$$\frac{1}{4 \times 63{,}360} = \frac{1}{253{,}440}$$

The U.S. Geological Survey uses somewhat of a



A sketch map and its graphic scales.

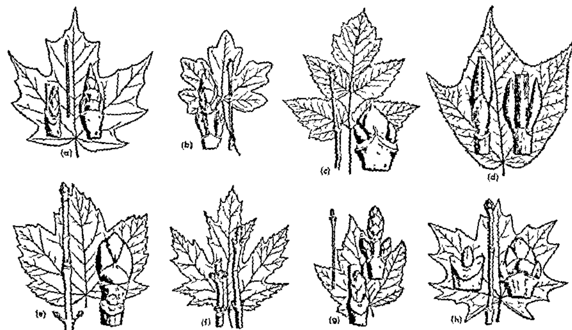
compromise between metric and British systems with such scales as 1:24,000, 1:62,500, 1:125,000, and 1:250,000 for topographic sheets. Countries or continents are shown at small scale, cities and small areas on large-scale maps. [E.RZ.]

Maple

A genus, *Acer*, of broad-leaved, deciduous trees including about 115 species in North America, Asia, Europe, and North Africa. This genus is characterized by simple, opposite, usually palmately lobed (rarely pinnate) leaves, generally inconspicuous flowers, and a fruit consisting of two long-winged samaras or keys. The winter buds have several overlapping scales, rarely only two. The most important commercial species is the sugar or rock maple, *A. saccharum*, called hard maple in the lumber market. This tree, attaining a height of 120 ft, grows in the eastern half of the United States and adjacent Canada. It can be recognized by its gray, furrowed bark, sharp-pointed scaly winter buds, and symmetrical oval outline of the crown. Maples rank third in the production of hardwood lumber. Michigan contributes the largest amount, about one-fourth of the 500,000,000 board ft cut annually. Hard maple is used for flooring, furniture, boxes, crates, woodenware, spools, bobbins, motor vehicle parts, veneer, railroad ties, and pulpwood. It is the source of maple sugar and syrup and is planted as a shade tree.

The red maple, *A. rubrum*, with much the same botanical range as *A. saccharum*, is second commercially to sugar maple. It can be distinguished by fewer scales on the winter buds, scalier or flakier bark, and the red coloration of all parts of the shoot, leaves, flowers, fruits, and twigs. The leaf lobes are separated by sharp sinuses in contrast to the U-shaped ones of sugar maple. Because it often grows in moist places, it is commonly known as swamp maple. In the lumber trade it is called soft maple.

The silver maple, *A. saccharinum*, with a range similar to *A. rubrum*, has more deeply cut leaves which are silvery beneath. Its lumber and that of the red maple are used for furniture, boxes, crates, woodenware, spools, bobbins, railroad ties, and pulpwood. Both species are planted as shade trees.



(a) Sugar maple, *Acer saccharum*. (b) Hedge maple, *A. campestre*. (c) Boxelder, *A. negundo*. (d) Striped maple, *A. pensylvanicum*. (e) Sycamore maple, *A. pseudoplatanus*. (f) Silver maple, *A. saccharinum*. (g) Red maple, *A. rubrum*. (h) Norway maple, *A. platanoides*. (A. H. Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

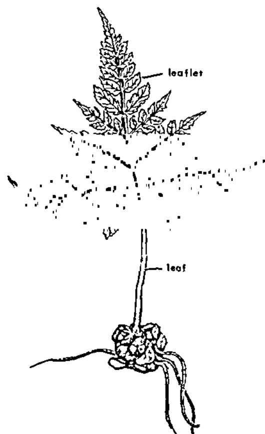
The bigleaf maple, *A. macrophyllum*, which attains a height of 90 ft in the Pacific Coast region and north to Canada, can be recognized by its large leaves which strongly resemble those of the eastern sugar maple, and by its long (about 1 ft) leaf stalks. The lumber is used for veneer, furniture, handles, fixtures, and woodenware. It is also planted as a shade tree.

The boxelder, *A. negundo*, with pinnately compound leaves, is native throughout most of the United States and part of Canada. It is often cultivated as a shade tree.

Two European species are popular street and shade trees: *A. platanoides*, the Norway maple, now becoming naturalized in North America, and *A. pseudoplatanus*, the sycamore maple. The former has copious milky sap in its leaves, twigs, and red bud scales, and close, nonscaly bark; the latter has green winter buds and bark which exfoliates in small polygonal plates. See FOREST AND FORESTRY; TREE. [A.H.C.]

Marattiales

An ancient order of ferns which in some respects is intermediate between the orders Ophioglossales and Filicales. Whereas their fossil remains indicate they were formerly world-wide and abundant, they now comprise only 7 genera and about 150 species limited to the humid tropical forests (see PALEOBOTANY). The highest concentration of genera and species is in southeastern Asia. They are mostly large ferns with persistent leaf bases. They resemble the Ophioglossales in habit and in having a massive eusporangiate type of sporangia (spore sacs develop from groups of epidermal cells). They resemble the true ferns in having the sporangia in sori on the lower side of the circinate (coiled) leaves. See FILICINEAE; OPHIOGLOSSALES; PTEROPSIDA. [P.A.V.]



Marattia alata Smith. Young sporophyte from which all but one of the leaves have been removed. (From G. M. Smith, Cryptogamic Botany, vol. 2, 2d ed., McGraw-Hill, 1955)

Marble

A term applied commercially to any limestone or dolomite taking polish. Marble is extensively used for building and ornamental purposes. See LIMESTONE; STONE AND STONE PRODUCTS.

In petrography the term marble is applied to metamorphic rocks composed of recrystallized calcite or dolomite. Schistosity, often controlled by the original bedding, is usually weak except in impure micaceous or tremolite-bearing types. Calcite (marble) deforms readily by plastic flow even at low temperatures Therefore, granulation is rare, and, instead of schistosity, there develops a flow structure characterized by elongation and bending of the grains concomitant with a strong development of twin lamellae. In calcite the twinning plane is the flat rhombohedron $\{01\bar{1}2\}$; in dolomite, twinning is markedly less common and the plane is the steep rhombohedron $\{02\bar{2}1\}$. See METAMORPHIC ROCKS; MINERALOGY.

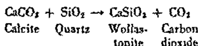
Regional metamorphic marbles are more or less deformed and therefore texturally marked by characteristic deformation patterns—parallel elongation of irregularly bounded lensoid grains of calcite (less typically of dolomite)—and by preferred orientation of mica flakes, if mica is present, producing a tightly woven, close fabric. This makes the rock, particularly the pure calcite (not dolomitic) marble, well suited for building purposes.

Contact metamorphism of limestones and of dolomites produces granoblastic rocks composed mainly of a mosaic of equant grains of calcite. The fabric is loose and the rock not suitable for use as building material.

Pure marbles attaining 99% calcium carbonate, $CaCO_3$, are often formed by simple recrystallization of sedimentary limestone. Dolomite marbles are usually formed by metasomatism. More interesting petrographically are impure marbles in which various reactions between silicates and carbonates have taken place.

If an impure carbonate rock is subjected to slowly increasing temperature at constant pressure, a series of reactions involving progressive elimination of carbon dioxide takes place. Each of these reactions is determined by the temperature, the pressure, and the bulk composition of the reacting system. In natural rocks, water and traces of other fugitive compounds are usually present. These facilitate the metamorphism and permit certain reactions to take place at low temperatures. The pressure usually has no great influence.

If at the beginning of the *reaction* there is a *dolomitic* limestone at low temperature composed of the phases *calcite* + dolomite + quartz (+ traces of *water*), certain mineral sequences will develop at given stages in the *metamorphism*. If the temperature is raised only a moderate amount, talc will form, followed by tremolite, forsterite, diopside, and brucite, until eventually wollastonite forms according to the following relation:



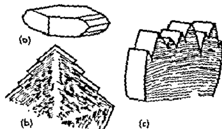
$$\underset{\text{Calcite}}{CaCO_3} + \underset{\text{Quartz}}{SiO_2} \rightarrow \underset{\text{Wollastonite}}{CaSiO_3} + \underset{\text{Carbon dioxide}}{CO_2}$$

This reaction has been carefully studied, and it can be shown that it takes place at about 450°C (assuming no great buildup of CO_2 pressure). At higher temperatures rare lime-silicates develop such as monticellite, akermanite, tilleyite, spurrite, merwinite, and others.

Examinations of natural occurrences show that at the relatively low temperatures of regional metamorphism the diopside stage is rarely passed. At the somewhat higher temperatures of the contact zones about *granites*, which come from the coolest of magmas, the wollastonite stage may be attained but not passed. It is only at the hotter contacts of syenitic or granodioritic masses that members higher in the series begin to form, and the highest members are associated only with basic, and usually basaltic, rocks. [T.F.W.B.]

Marcasite

A mineral having composition FeS_2 and crystallizing in the orthorhombic system. Marcasite crystals are common, usually tabular parallel to the basal plane, and characteristically twinned, giving cockscomb groups (see *illustration*). It frequently has a radiating structure and may be globular or stalactitic. There is poor prismatic cleavage. The hardness is 6–6.5 (Mohs scale) and the specific gravity is 4.89. The luster is metallic and the color pale bronze yellow to nearly white on a fresh fracture. Marcasite and pyrite are dimorphous; both have the composition FeS_2. Because marcasite is whiter it is called white iron pyrites. It is much less common and less stable than *pyrite*. Specimens may in a year or two completely disintegrate, with the formation of ferrous sulfate and sulfuric acid.



Marcasite crystals. (a) Orthorhombic, dipyramidal crystals in prisms. (b-c) Twinned cockscomb crystals. (From C. S. Hurlbut, Jr., Dana's Manual of Mineralogy, 16th ed., Wiley, 1952)

Marcasite is usually found in surface or near-surface deposits in which it was formed at low temperatures from *acid* solutions. Most marcasite is supergene but in some places it is believed to have been deposited near the surface from ascending vein solutions. Under these conditions it is usually the last mineral to be deposited.

It is found in metalliferous deposits associated with lead and zinc ores, as replacement deposits in limestone, and in concretions in clays and shales. The nodular and lenticular masses in coal known as brasses are in part marcasite and in part pyrite. *See* PYRITE. [C.S.HU.]

Margarine

A plastic food fat product composed of processed vegetable oils or animal fats or both, cultured milk, salt, and emulsifiers. Margarine was invented by Mége-Mouriés, who in 1870 received a prize offered by Louis Napoleon for a satisfactory butter substitute. It has been improved so that its flavor is almost indistinguishable from that of butter costing about two and a half times as much as margarine.

Manufacture. Pasteurized whole or skim *milk* with 0.1–0.2% added citric acid is ripened in an agitated vessel at approximately 60°F. Ripening is induced by inoculation with 2–3% of a starter containing mixed cultures of streptococci selected for their ability to produce desired flavors (*see* BUTTER). After ripening about 16 hours, the cultured milk is cooled to 35–40°F, salt is added, and the milk is blended with margarine oil containing (as *emulsifiers*) lecithin or mono- and diglycerides (or all three). A water-in-oil emulsion is formed, and this can be carried out by churning or by chill rolls. In American practice the continuous votator process is used almost exclusively (Fig. 1).

Votators are efficient heat-exchange units for chilling and mechanically working a continuous stream of emulsion (Fig. 2).

The solidified margarine is formed into prints and wrapped after it emerges from the votators.

In the United States, hydrogenated soybean or cottonseed oils are usually used. In other countries peanut, rapeseed, sunflower, coconut, palm, palm kernel, and whale oils and animal fats are employed. Margarine's plastic properties are controlled by blending of hard and soft oils or by hydrogenation. *See* COTTON; HYDROGENATION; PEANUT; SOYBEAN.

Laws in various states, provinces, and countries regulate addition of vitamin A, coloring, flavoring, antioxidants, preservatives, emulsifiers, and indicating agents such as starch or sesame oil, and in some cases the types of oil to be used.

Margarine contains not less than 80% fat, 1–3% salt, 0.3–1.0% milk solids, up to 0.5% emulsifiers, and usually 15,000 USP units of vitamin A per pound. *See* VITAMIN A.

Production. Production in the United States for the 5 years 1954–1958 averaged 1,400,000,000 lb. It exceeded 1,500,000,000 lb in 1958, utilizing 1,000,000,000 lb of soybean oil (which was 84% fats and oils used) and 145,000,000 lb of

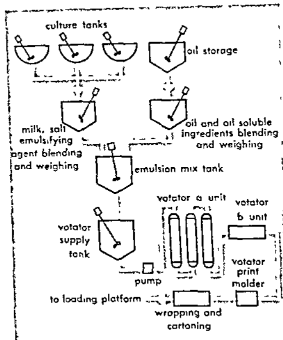


Fig. 1 Flow diagram for continuous margarine manufacture. (*J. E. Slaughter, Jr., and C. E. Michael, J. Am. Oil Chemist's Soc., 26 623–628, 1949*)

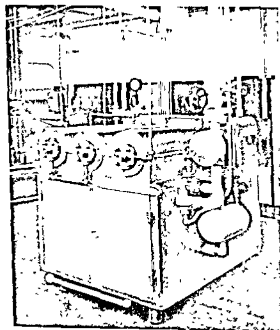


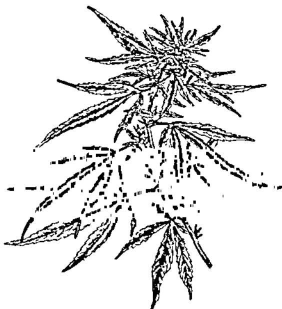
Fig. 2. Margarine votator unit. (*Girdler Process Equipment Division, National Cylinder Gas Co.*)

seed oil. Per capita consumption has risen from 2.9 lb in 1935 to 9.0 lb in 1958. World production is over 6,000,000,000 lb. *See* FAT AND OIL, EDIBLE; FOOD ENGINEERING. [F.C.D.]

Bibliography: A. J. C. Andersen, *Margarine*, 1954; A. E. Bailey, *Industrial Oil and Fat Products*, 2d ed., 1951; U.S. Dept. Agr., AMS, *The Fats and Oils Situation*, FOS-196, 1959.

Marijuana

The Spanish name for the dried leaves and flowering tops of the hemp plant, *Cannabis sativa*, which are illicitly introduced into cigarettes. The narcotic ingredients allegedly have stimulating effects and after smoking two or three cigarettes, the smoker experiences a feeling of well-being and increased power and ability. Illusions are often common, as well as pleasing, fanciful hallucinations. Sometimes the addict experiences disorientation and even delirium. The influence of marijuana has induced individuals to commit violent deeds including murder, sexual orgies, and gruesome sadistic acts. The plant thrives as a weed in waste places and is surreptitiously grown in window boxes; consequently material for smoking is inexpensive and can be easily obtained. Despite sharp legal restrictions, the use of marijuana has not been eliminated. *See* URTICALES. [P.D.S.]



Marijuana (*Cannabis sativa*). (*From H. W. Youngken, Textbook of Pharmacognosy, 5th ed., McGraw-Hill, 1946*)

Marine boiler

A steam boiler designed to meet requirements of ship operation (*see* STEAM BOILER). Steam generated in marine boilers may be used as the working medium in (1) a prime mover for ship propulsion, (2) auxiliary machinery, or (3) generation of electricity. Steam is also used to heat passenger, crew, or cargo spaces.

Types of units. There are two important types of boilers in marine use—the fire-tube and the water-tube. The Scotch marine boiler is the marine adaptation of a fire-tube boiler. Although still used, the trend to higher steam pressures and temperatures has greatly limited its present application. *See* FIRE-TUBE BOILER; WATER-TUBE BOILER.

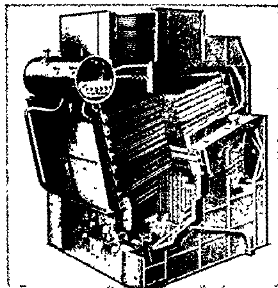


Fig. 1. Header-type marine boiler, fitted with superheater and air heater and arranged for oil firing. (Babcock and Wilcox Co.)

There are two types of marine water-tube boilers—the header type and the drum type. In marine header-type boilers (Fig. 1) the generating tubes are straight and expanded into vertical sinuous headers. These generating sections are in turn connected at the top to a horizontal steam drum. The headers acting as downcomers are connected to a horizontal mud drum. In marine drum-type boilers there are various arrangements of tubes, drums, and furnaces which follow certain patterns as shown in Figs. 2 and 3. The two-drum type (Fig. 2) is used most extensively.

Special requirements. In the design of marine boilers certain special requirements must be considered. The units must be compact and economical of space and weight. They must be built to withstand the effect of pitching and rolling of ships at sea as well as the vibration and shock encountered during such operation. Reliability must be inherent in the design because of the requirements of safety at sea.

Design consideration and fuels. Marine boilers are generally arranged for oil firing. Oil is used extensively because of the simplicity of handling and storing the liquid fuel and the ease with which the firing rate can be changed during maneuvering. For economic reasons coal is used in certain regions, as on the Great Lakes. Wood or by-products may also be used.

Oil-firing boilers are arranged for forced draft; that is, the air for combustion is blown or forced into the boiler furnace. Coal firing uses natural, forced, or induced draft or a combination of these arrangements. Space limitations require marine boilers designed to burn large quantities of fuel in small furnaces.

The oil normally used is a cracked residual fuel called bunker C, heated to the proper viscosity and pumped under pressure to the burner. It is atomized mechanically or with steam or air.

In coal-burning installations, coal is burned on a grate and fired by hand or by mechanical stokers. The spreader-type stoker has been installed in most cases.

Marine boilers are fitted with superheaters to increase steam temperature and with heat recovery apparatus to obtain increased efficiency.

Automatic controls. While many marine installations are hand controlled, automatic controls are becoming important. Controls are applied to regulate steam pressure, steam temperature, and feedwater, air, and fuel inputs. Generally, automatic controls are either electrically or air operated. Mechanically and thermally operated boiler-water level controls have also been installed.

Feed-water flow to the boiler is regulated by suitable valves and pump controls in accordance with the water level in the drum, steam flow from the boiler, water flow to the boiler, or combinations of these elements.

General performance. Reliable and efficient performance is required over the economic life of the ship, currently considered to be about 20 years. Normal maintenance and repairs are made while a ship is in port. Major repairs are made in shipyards during the period of the annual inspection required by law.

Boiler firesides must be clean of soot and ash deposits to obtain high boiler efficiency. For this purpose soot blowers are installed and the boiler is water-washed periodically.

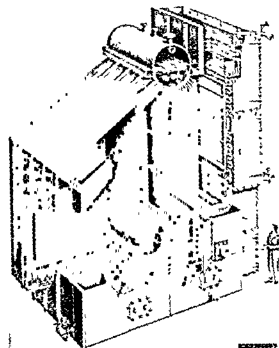


Fig. 2. Two-drum type marine boiler fitted with superheater and economizer and arranged for oil firing (Babcock and Wilcox Co.)

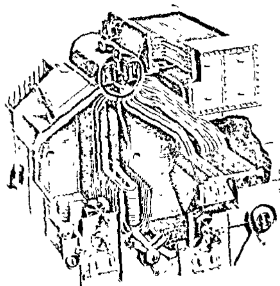


Fig. 3. Single-uptake controlled superheat marine boiler fitted with superheater and economizer and arranged for oil firing (Babcock and Wilcox Co.)

Scale deposits on the water side are prevented by proper water conditioning (see BOILER FEED WATER). High heat-liberation rates in marine boilers resulting from compact installations make this particularly important. Chemically treated distilled water is practically mandatory for this boiler.

Most modern American merchant ships utilize steam at a pressure of 600 psi or 850 psi with steam at 850°F, but present trends indicate an increase in steam temperature. For naval use, conditions of 1200 psi and 950°F are currently used. See MARINE MACHINERY. [C.H.H.]

Marine ecosystem

The ocean with its shores and estuaries is the largest conceptual unit in marine ecology. Within it, ecological systems of various sizes are recognized, for example, the particular seaweed ecosystems, tidepool ecosystems of the seashore, and estuarine-bay ecosystems (see COMMUNITY). Thus small, natural communities of organisms with their immediate environment, the unit marine ecosystems in the narrow sense, exist within and together compose larger systems. The size limits, or boundaries, of a particular ecosystem that distinguish it from others may depend upon physical barriers to community dispersal, like a submarine mountain; environmental factors that restrict the area of the biotic community, such as the salinity factor; the productivity cycles within the community; or the reproductive and dispersal potentials of the community. This conceptual unit of ecological science, whether large or small, possesses individuality, a degree of stability and permanence, characteristic functional cycles, and readily recognizable components, either living or nonliving or both. See ECOSYSTEM.

The nonliving, or abiotic, materials in a marine ecosystem cycle comprise not only a variety of water-soluble inorganic nutrient salts, such as the phosphates, nitrates, and sulfates of calcium, potassium, and sodium and the dissolved gases oxygen and carbon dioxide, but also organic compounds, like the various amino acids, vitamins, and growth substances. Most of the solid material dissolved in the sea originated from the weathering of the crust of the earth.

MARINE ENVIRONMENT

The marine environment is a subject of study by oceanographers who investigate the physical, chemical, and biological properties of ocean waters, ocean currents, and ocean basins. See ENVIRONMENT; OCEANOGRAPHY.

Chemical factors. The chemical properties of ocean waters comprise an important aspect of the marine environment. Such chemical factors as the dissolved solids and gases in the ocean waters have been investigated. See SEA WATER.

Dissolved solids. The total quantity of dissolved solids in the waters of the world's oceans or in the marine ecosystem as a whole approximates 5×10^{16} metric tons, enough to form a layer 153 in. thick over the land area of the earth. Typical ocean water contains about 34.9 grams (g) per liter of dissolved materials, in which there are 19.3 g of chlorine, 10.7 g of sodium, 2.69 g of sulfate, and 1.31 g of magnesium. Some of the most important nutrient elements are present in very small amounts, for example, the phosphates needed for the growth of algae. Phosphorus exists in ocean waters as phosphate ions. It is an essential component of living things, and in ocean water the amount present may limit the production of plants. The inorganic form occurs in amounts varying from zero concentration



Fig. 1. A ribbed-mussel community of a decayed log in the intertidal zone.



Fig. 2. A marsh grass-mussel community representing a small marine ecosystem of the intertidal zone

to over 0.10 mg per liter. Phytoplankters, principally microscopic algae, may absorb inorganic phosphorus and reduce the amount remaining in the water to a minimum. Some is permanently lost to recycling by being bound in nonsoluble forms and deposited on the sea bottom. By their death and decay, phosphorus is returned to the aquatic environment. Certain of the 40 or more elements in sea water which, like phosphorus, exist in extremely small amounts are concentrated to important degrees. For example, macroscopic algae concentrate potassium and iodine in relatively large amounts. Of special importance is silicon, utilized by diatoms and other silica-secreting organisms.

Dissolved gases. The essential gas, carbon dioxide, is readily dissolved in sea water, but in equilibrium with the air it would contain only about 0.5 milliliters (ml) of free CO_2 per liter at 0°C. Actually, sea water contains about 47 ml of CO_2 per liter because appreciable amounts are present in the form of carbonate and bicarbonate ions. This gives great importance to the marine ecosystem as the world's reservoir of CO_2, although only about 1% is in the free form. Since, in relation to the atmosphere, its CO_2 content is 50 times greater, oceans regulate the concentration in the atmosphere. The CO_2 content of sea water originates not only from the atmosphere but also from biochemical processes in the sea—respiration and decay—and in the soil of the ocean bottom and coastal shores. The free carbon dioxide in water exists in simple solution and in the form of carbonic acid, H_2CO_3. Combined CO_2 is in the bicarbonate ions (HCO_3^-) and the carbonate ions (CO_3^{--}). These and other ions of weak acids constitute the buffer system of sea water that stabilizes its hydrogen-ion concentration, or pH, at about 8–8.4 at the surface and at 7.4–7.9 at the deeper levels. The high buffer capacity of sea water is a dominant environmental characteristic. Chemical and biochemical interactions and actions of biological agents which would otherwise produce pH changes that could seriously modify living conditions are largely negated by this buffer system. The ocean environment is thus chemically stable. Marine plants, the producers, may utilize primarily either free CO_2, or some free CO_2 and some combined CO_2. In any case, lack of CO_2 is not a limiting factor to growth of ecosystems in the sea. High concentrations, however, are a limiting factor to many marine animals such as fish.

The dissolved oxygen content in the sea-water environment varies from zero to over 8.5 ml per liter. The oxygen in water comes from the atmosphere by diffusion, and from aquatic plants by their photosynthetic action. Low temperatures and low salinities of water favor high solubilities of oxygen gas. There are low-oxygen strata in ocean waters, but in general the supply is adequate for animals. The considerable oxygen content of the deeper water layers of the ocean reached these layers when they were in the photosynthetic surface strata.

Physical factors. The physical aspects of marine ecological systems, temperature and salinity, are especially significant because of their degree of constancy away from land. The mean annual temperatures in different latitudes on the earth remain unchanged. Seasonal variations in the ocean are small compared to those on land. Also, the entire temperature range in ocean waters is within the tolerance limits of numerous plants and animals. In polar and tropical seas, temperatures do not vary more



Fig. 3. Plankton diatoms, with light spiny shells suitable for floating (Photograph by P. Conger, Smithsonian Institution)

than about 5C° during the year. Temperate seas commonly vary about 10–15C°. The extent of seasonal variations decreases in the deeper strata. In temperate and tropical regions a permanent ther- [illegible] high temperature [illegible] sur- [illegible]. Low temperatures, around 3°C, in the deep and bottom waters of the ocean exist because the waters of greatest density, formed in high latitudes, sink to *the bottom or to levels of similar density,* then spread out and move toward warmer latitudes, to form a pattern of oceanic circulation. The salinity of the world's oceans varies only a few parts per thousand and averages around 35‰. Major latitudinal differences in density result from temperature differences. As expected, there is a vast system of oceanic circulation involving all depths and modified by large water masses contributed by adjoining seas. *Thus, this offshore environment* of the unit marine ecosystem is characterized by relative constancy of salinity over exceedingly large areas. The massive and relatively constant properties of this environment are reflected in the distribution, abundance, *and morphological traits* of its biota.

Biota. The organisms that comprise the biota of the marine ecosystem are characterized by a lack of diversity among the plants, in contrast to the animals. *This generalization* applies to both microscopic and macroscopic organisms. With few exceptions, all types of animals inhabiting land and fresh-water environments occur in the marine ecosystem. Six major animal groups, including ctenophores, starfishes, and certain worms, *are restricted* to the marine environment. Other major taxonomic units are predominantly marine. The kinds of plants *and animals* in the marine environment fall into three major groups, organisms of the plankton, nekton, and benthos.

Plankton organisms. These organisms are small, mostly microscopic, and have little or no power of locomotion, being distributed by water movements. There are two main types, the phytoplankton and zooplankton. The former includes all of the floating plants, such as the small algae, fungi, and sargassum weeds. Of these, the *most important* in the economy of the sea are algae—diatoms and dinoflagellates. They are the major producers in marine plankton (Fig. 3). Diatoms are microscopic, unicellular plants, *some of* which form chains. They possess characteristic shells composed of translucent silica, and have a great variety of form and sculpture. The shell structure consists of two nearly equal valves, one of which fits over the other and hence may be compared to a box with a telescoping lid. The valves are joined by connecting bands. The protoplasm within the shell is exposed by a slit or by small pores to permit *metabolic* interchanges with the environment. During reproductive division of the protoplasm of the diatom, one of the two protoplasmic daughter cells retains the larger epivalve, or lid valve, the other the hypovalve, *or box valve.* The daughter cells then lay down the needed complementary valves. This simple binary form of division is the most common one. It permits rapid production of vast populations in favorable environments of an ecosystem. During successive binary divisions, the size attainable by *individuals* is *progressively* reduced until a minimum limit is reached when, usually, the protoplasmic content of the shells escapes from the parted valves. It is enclosed in a flexible pectin membrane and is called an auxospore. These *specialized spores grow in size* and finally form the characteristic two valves. Diatoms possess one or more chromatophores, ranging in color from yellow to brown. *See* DIATOM.

Diatoms occur as *fossil, siliceous* shell deposits, called diatomaceous earth, and as living producers in practically all habitats of the broad marine ecosystem. They are found floating in water, attached to the bottom, on larger plants, on animals, and, as [illegible] diatoms [illegible] ments in depth. The bladder-type [illegible] tively larger, and in some forms such as *Planktoniella*, the shape is disklike, so that a *zigzag course is followed in sinking. The hair type,* such as *Rhizosolenia*, is a long and slender diatom, which sinks slowly when the long axis is horizontal to the pull of gravity, but more rapidly when oriented vertically. The ribbon-type cells such as *Fragillaria are broad and flat, and are attached* to form chains.



Fig 4. *Chaetoceros atlanticus* Cleve, branched type, oceanic diatom. (*Photograph by P. Conger, Smithsonian Institution*)

The most abundant diatoms in the off-shore oceanic waters are of the branched type, such as *Chaetoceros*. They possess numerous spiny projections that resist sinking (Fig. 4).

Dinoflagellates possess whiplike flagella that provide a slight degree of locomotion. Like diatoms, they possess structural modifications that indicate adaptation to environmental conditions. Some, such as *Dinophysis*, possess winglike structures that favor suspension; some have cellulose plates; others are naked cells. Many kinds are luminescent. *See* BIOLUMINESCENCE.

Phytoplankton organisms are much more abundant in nutrient-rich coastal waters than in offshore oceanic waters. They are the primary producers upon which large and small marine animals feed.

Zooplankton organisms are the floating or weakly swimming animals, which include the eggs and larval stages as well as adult forms. The principal kinds include numerous Protozoa, such as Foraminifera and Radiolaria; great numbers of small Crustacea, such as ostracods and copepods, with their various larval stages; various jellyfishes; numerous worms; a few mollusks; and also the eggs and early developmental stages of most of the nonplanktonic organisms in the sea. Plankton organisms are grouped on a basis of size. The smallest organisms range from about 5 to 60 microns, the next largest range up to about 1 mm in size, the next up to 1 cm, and the largest plankton group over 1 cm.

Nekton organisms. These are the actively swimming animals of the marine ecosystem. They comprise adult stages of such familiar forms as crabs, squids, fish, and whales. Some of these undergo long horizontal migrations over hundreds of miles; some migrate periodically to great depths; and a few live mostly in deep waters, for which habitat they possess marked adaptations.

Benthos organisms. These inhabit the bottom, and range from high-tide level on shore to the deep-sea bottom. There are relatively few kinds and numbers of animals on the deep-sea floor. They are mainly mud dwellers, possessing characteristic structures permitting life in a quiet, dark, muddy environment where food is scarce. Certain isopods, sponges, hydroids, brittle stars, sea urchins, and shrimp are typical animals of the deep-sea biota (Fig. 6). Inshore bottom communities at depths less than 50 meters (m) are rich in plants and animals. Although mosses and ferns are entirely absent in the sea, there are approximately 30 species of marine flowering plants. Of these, eel grass, *Zostera*, is a characteristic shore plant. It is a perennial flowering plant, not actually a grass, that is world-wide in distribution, and is most abundant on soft bottoms in protected, coastal areas. Large brown algae, such as *Fucus* and *Ascophyllum*, are widely distributed, conspicuous plants of exposed rock surfaces in the intertidal zone. *Ulva* and *Enteromorpha* are typical green algae of mud flats in quiet waters. Red algae, such as dulse, *Rhodymenia*, are commonly found below the intertidal zone. These are important food plants for the bottom-living animals of the coasts.

Conspicuous members of the benthic animal communities of the sea coast are barnacles, snails, mussels, clams, oysters, sea anemones, sea urchins, sea cucumbers, and starfish. Shore corals are largely restricted to warmer seas. The bottom animals of the marine ecosystem have structural modifications facilitating adhesion, burrowing, feeding, and protection.

MAJOR DIVISIONS OF THE MARINE ECOSYSTEM

The marine ecosystem can be divided into two large areas, the pelagic and benthic divisions. Each of these consists of various zones.

Pelagic division. This division embodies all the waters of the oceans and their adjacent salt-water bodies. It is divisible into the neritic zone, extending offshore to the edge of the continental shelf to a depth of over 200 m, and the oceanic zone, embracing the remaining offshore waters. The neritic zone is rich in plant nutrients, especially phosphates and nitrates, originating from coastal tributary waters and from bottom deposits carried upward to surface water layers by upwelling, diffusion, turbulence, or convection. The water is more variable in density and in chemical content than oceanic waters. It is also far more productive of plankton, fish, and shellfish. Many of the inshore, coastal forms are adapted to withstanding brackish waters of coastal tributaries. The oceanic zone has a well-populated upper, lighted, 200-m stratum, and deeper, relatively dark, and sparsely populated layers, characterized by great pressures, animals modified for life in darkness and under great pressures, and very few bottom animals. Since there is less plankton and suspended organic material in the water, light penetrates further than in the neritic zone. Also the water is usually very transparent. Its salt content is uniformly high and not subject to major variations in time and space. In the upper photosynthetic zone, or zone of productivity, plant nutrients are much less concentrated, and the cycle for replenishment is much longer than in coastal waters that receive nutrients and organic detritus from the land.

Benthic division. This division of the marine environment embraces the entire ocean floor, both coastal bottom and deep-sea bottom, properly termed the littoral and deep-sea systems, respectively. The littoral system consists of the eulittoral zone and the sublittoral zone.

Eulittoral zone. This extends from high-tide level to about 60 m, the approximate depth below which attached plants do not grow abundantly. This is probably the richest zone of the marine ecosystem in respect to numbers and kinds of organisms, as well as in variety of ecological types and habitat modifications. The upper intertidal portion of the zone extends between high- and low-water marks, a vertical distance that varies on different continental shores from over 12 m to a few centimeters. Changes in such environmental factors as light, temperature, salinity, and time of exposure vary tremendously within short vertical distances across the zone. These variations are reflected in th

Fig. 5. A littoral community with abundant sand dollars. Mission Bay Channel, San Diego, California. Depth 4 m. (Photograph by R. F. Dill, U.S. Naval Electronics Laboratory)

shapes, movements, tolerances, and life histories of the characteristic animals and plants. Numerous small, partly independent ecosystems thrive in this area because the substratum includes a variety of rock exposures, gravel, sand, and mud types admixed in all degrees (Fig. 5). The lower, permanently submerged portion of the eulittoral zone is characterized by abundant sessile plants, such as conspicuous rock weeds (*Fucus*), bladder-wrack (*Ascophyllum*), green sea lettuce (*Ulva*), and kelp (*Laminaria* and *Nereocystis*), many of which are common to both portions of the zone. Certain of those algae, such as the giant kelp, form productive algal forests that extend down below the typical eulittoral zone into the sublittoral zone. Here, they utilize the dimly lighted, lower, and least productive portion of the littoral system.

In tropical littoral waters, the coral reef communities abound with characteristic types and forms of both plants and animals. Visible algal growth is sparse, animals often greatly predominate over plants, and depth range is to about 60 m; hence the communities occur in both zones of the littoral system. The littoral system terminates at depths varying in different latitudes between about 200 and 400 m, depending upon light and temperature factors that modify the distribution of benthic animals and plants.

Deep-sea system. This portion of the benthic division is subdivided into an upper part, the archibenthic zone or, more meaningfully, the continental deep-sea zone, extending from the edge of the continental shelf (200–400 m) to depths of about 800–1100 m, and the abyssal-benthic zone that embraces the remainder of the benthic deep-sea system (Fig. 6). These zones have little or no light, relatively constant conditions of salinity and temperature, and steadily decreasing numbers and kinds of organisms. In the abyssal regions of great depth, perpetual darkness and extremely low temperature (5 to −1°C), the extreme in monotony of environment [illegible]
some [illegible]
remai[illegible]
contin[illegible] fall down to settle over the bottom as a light coating, utilized by the bottom dwellers for food. The pelagic food supply from above decreases with the increase in offshore distance for two reasons: the offshore oceanic waters contain less particulate matter than inshore waters, and the longer period required to descend the deeper strata results in greater disintegration while sinking. *See* DEEP-SEA FAUNA.

Estuaries. These are coastal adjuncts of the marine ecosystem. They embrace bodies of water which, by virtue of their position, are directly subject to the combined action of river and tidal currents. Compared with offshore ocean waters, they lack constancy, possessing characteristic horizontal and vertical gradients in physical, chemical, and biological properties. These gradients are subject to pronounced changes in space and time. Estuarine waters are characterized by rapidity of response to changing external conditions. Their temperature, salinity, and turbidity conditions are distinctly not uniform. They are usually rich in plant nutrients of land origin, which they transport to coastal waters and thus fertilize these waters. Estuarine waters contain an environmentally selected biota containing representative fresh-water and salt-water forms. *See* ESTUARINE OCEANOGRAPHY.

Marshes. These transitional land-water areas, covered part of the time at least by estuarine or coastal waters, comprise parts of the peripheral area of the marine ecosystem. Mangroves are characteristic woody marshes in tropical tidal waters of flat, muddy shores. More generally, coastal marshes are dominated by grasses and sedges. Marshes have characteristic biota for a certain latitude. In temperate regions, they are inhabited by characteristic birds such as bitterns and rails, and by crustaceans such as sand hoppers and fiddler crabs. Typical marsh animals and plants have tolerances to fresh water and salt water that differ from those of related forms living in fresh-water or salt-water habitats. *See* MANGROVE SWAMP; MARINE MARSH.

Sediments. Inanimate, particulate materials of organic and inorganic origin that have settled on

Fig. 6. A deep-sea benthic community of brittle stars and isopods, depth of 1200 m in the San Diego Trough, 32°54'N and 117°36'W. (Photograph by G. A. Shumway, U.S. Naval Electronics Laboratory)

the bottom in aquatic environments comprise the sediments. Vast quantities of particulates, eroded from land surfaces by natural waters, are carried to the sea by rivers and estuarine waters. They settle to the bottom and provide food for benthic animals. Sediments may be carried to the surface, or photosynthetic stratum, either by upwelling of coastal waters or by ocean currents, and thus enter a biochemical cycle of the marine ecosystem. Marine sediments are comprised of microscopic fragments of weathered rock, partly decomposed plant and animal remains, skeletal remains of organisms, inorganic precipitates from sea water, terrestrial particulates, and particulate material of volcanic origin. See MARINE SEDIMENTS. [C.L.N.]

Bibliography: H. B. Moore, *Marine Ecology*, 1958; C. L. Newcombe, Mussels, *Turtox News*, vol. 25, no. 1, 1947; E. P. Odum, *Fundamentals of Ecology*, 1953; H. U. Sverdrup, M. W. Johnson, and R. H. Fleming, *The Oceans, Their Physics, Chemistry, and General Biology*, 1942.

Marine engine

An engine for propelling a water-borne vessel. Even in small craft the marine engine must have the following characteristics: reliability, light weight, compactness, fuel economy, low maintenance, long life, relative simplicity for operating personnel, ability to reverse, and ability to operate steadily at low or cruising speed. The relative importance of these characteristics will vary with the service performed by the vessel, but reliability is of prime importance. See MARINE MACHINERY; SHIP PROPULSION.

Steam engines. Steam, used to drive the earliest powered vessels, is still the favorite type of propulsion for large ships throughout the world. The diesel engine has gained wide acceptance in foreign merchant ships of medium size, but in the United States steam propulsion remains the majority choice for all but the smallest vessels.

Reciprocating steam engines. Early engines commonly used steam in series flowing through as many as four cylinders whose pistons had the same stroke but were of increasing diameters. This system provided for an expansion or increase in steam volume which accompanied the decrease in pressure to the exhaust under a vacuum. The modern multicylinder uniflow marine steam engine, with complete expansion in each cylinder, shows better steam economy. Because it has the same diameter for all cylinders (2–6 in number) it is preferable from a manufacturing viewpoint. Equal power is developed by each cylinder; units of four cylinders or more have good torque and balance characteristics. A steam rate of 10 lb per indicated horsepower per hour with 275 psi, 240°F superheat is attained. Uniflow engines of up to 5000 hp have been fitted on shipboard. Normally, steam engines are double acting; that is, steam acts on each side of the pistons. With superheated steam, piston-cylinder lubrication must be provided. Pure feed water is required by modern high-capacity boilers; therefore an effective filter is fitted where the condensate must be used as feed water.

Steam turbines. The marine steam turbine has the advantages of direct rotary motion, little or no rubbing contact of pressure-confining surfaces, and ability to use effectively both highly superheated steam and steam at low pressure, that is, at a high vacuum where specific volumes of over 400 ft^3/lb are reached

For good efficiency of steam turbines, high rotative speeds are required. This requirement led to the introduction of the mechanical reduction geared turbine and turboelectric drive. These systems give efficient turbine speeds and efficient propeller rpm. With geared turbines, for example, turbine rotor speeds range from 3,000 to 10,000 rpm while propeller rpm is reduced to the 80–400 range.

In low-powered geared turbines, steam completes its expansion in one rotor and casing. Such a design has been used in geared turbines of up to 4000 shaft horsepower (shp). However, series flow through two, or even three, casings is preferable in most steam turbines. This arrangement provides for more flexibility in turbine design, allowing for different and optimum revolutions for high- and low-pressure rotors. Also, in a seagoing vessel in case of casualty to one turbine or its high-speed pinion, the vessel usually can make port with the remaining turbine in operation.

Multistage pressure compound (Rateau) type and reaction (Parsons) type turbines, the latter with a first-stage velocity compound (Curtis) wheel, are about equal in favor and economy. With the cross compound, that is, two casing turbines of the Rateau type, 6–10 ahead, pressure stages are fitted on each rotor. Reaction turbines have more stages, their blading being set in grooves of a drum-type rotor. To reduce the axial thrust that would be caused by the steam pressure at the turbine inlet, a dummy or balance cylinder, sealed by labyrinth packing, is fitted at the high-pressure end.

Astern operation is obtained by fitting an astern turbine, usually at the exhaust end of the low-pressure turbine. Such an astern turbine consists of a velocity compound two-blade row stage, often followed by a second velocity compound stage or even a single row stage. By this means, about 80% of full-ahead torque at half of full-ahead rpm is obtained for cargo ships and tankers. This is enough to stop and reverse the propeller and the ship. Because the astern turbine rotates in a vacuum during ahead operation, the windage or rotation loss due to it is of minor importance.

Internal combustion engines. Both diesel and gasoline internal combustion engines are used in marine applications. Many moderate- and low-power marine installations utilize automotive or locomotive engines designed for variable load and intermittent service. High-power marine propulsion units normally are called on to operate continuously under load. Therefore, the brake horsepower (bhp) rating of units selected for marine service should be conservative.

(a) (b)

Fig 1 (a) Cross section of 6-cylinder trunk piston Nordberg main diesel engine. (b) Cross section of power cylinder (Nordberg Mfg. Co.)

Diesel direct drive. For typical commercial freight vessels 300–600 ft in length, diesel direct drive provides economical service speeds up to 14 to 19 knots respectively. For good propeller efficiency, the propeller rpm should be under about 120. Such a top limit on engine revolutions results in a large, heavy, bulky, slow-rpm engine. However, the direct-drive diesels have a lower fuel oil consumption than do higher-rpm units, and with suitable fuel treatment will operate on the better grades of the cheaper fuel oil burned in boilers.

In the United States, two-cycle direct-drive single-acting diesel engines up to 8 cylinders and 6000 bhp have been fitted (see Fig. 1). There has been a tendency to operate at speeds above an efficient propeller speed in the interests of smaller engine size and lower weight. The Sun-Doxford opposed piston engine has been installed in similar powers; it is inherently of low rpm. However, for powers per screw of over 4000, the geared turbine is generally preferred in the United States.

Slow-speed diesel direct dr[illegible]

[illegible]

[illegible] bhp are now available; such engines weigh over 100 lb/bhp. The total machinery weight for powers over 2000 bhp is more than the weight of geared turbine machinery including boilers and auxiliaries.

Two-cycle single-acting turbo-supercharged diesel marine engines of over 17,000 bhp are under development. Tests of units designed for 10 or 12 cylinders of diameters up to 36 in. have been made. In the past, the diameter has been limited to 32 in. because of heat stresses.

Moderate-speed diesels. Diesel engines of 250–500 rpm are available in 2- and 4-cycle single-acting types, generally with trunk pistons. In some marine applications, they are connected directly to the propeller and thus fitted only with reverse gear. However, they are also employed with geared diesel and diesel-electric drive. The weight of such engines runs about 35–70 lb/bhp.

High-speed diesels. Many high-speed diesel engines (trunk piston-type) of 600 rpm and over (some of types originally developed for truck and locomotive service) are available for marine propulsion. Opposed piston types have been developed by Fairbanks, Morse and Company and by Napier (British Deltic); other manufacturers favor a "V" type to reduce weight. Such engines are of 2- and 4-cycle types and usually weigh 10–40 lb/bhp. Because of less efficient scavenging, lower powers, and other factors, their fuel and lubricating oil rates are higher than for large, low-speed diesels.

Except for direct drive in moderate or fairly high-speed craft, marine applications of diesel engines are fitted with either mechanical reduction gearing or with diesel-electric drive to provide good propeller efficiency. Because their pistons, valves, and other components are small, standardized, and carried in stock, repairs are readily made, if spare parts are available, with the result that engines of this type are popular for non-ocean-going services.

The greater weight of large 4-cycle engines, despite their slightly better fuel rate, and the great difficulty in repairing large double-acting diesel engines have decreased the demand for these two

types. The Doxford opposed piston engine and the Burmeister-Wain piston valve type, however, are often fitted in foreign vessels.

Oil consumption and starting. Lubricating oil consumption of diesel engines is high, because of the cylinder-piston lubrication that must be provided and the contamination of the crankcase oil with residues blown by the piston rings. In large engines, this contamination is avoided by using piston rod–crosshead construction so that crankshaft, connecting rods, and crossheads operate in a closed casing separated from the working cylinder

These engines are started and maneuvered by air pressure from one or more reservoirs filled with air at about 250 psi. To make it feasible to start and readily reverse, 2-cycle single-acting marine engines should have at least 4 cylinders; with 4-cycle engines, 5 or more cylinders should be fitted.

Mechanical reduction gears. Reduction gearing for diesel and gasoline engines allows the use of a relatively high engine speed and lower, more efficient propeller rpm. Speed reduction ratios of 1.8:1 to 4:1 are common, preferably with helical teeth to give better wear and quieter performance A reverse gear device often is incorporated in low-power gears for astern operation, or the engine may be made reversing. Another method of maneuvering involves the use of a controllable pitch propeller. See PROPELLER, MARINE.

One, two, three, or four engines may drive the same gear through individual pinions. The use of a friction, electromagnetic, pneumatic, or hydraulic coupling serves to disconnect any engine. By reversing one or more engines, ready maneuvering, including astern operation, is provided for by the use of the respective coupling.

The high rpm (3000–9000) of modern marine steam turbines and the low revolutions of an effective propeller (as low as 80 rpm) require the use of double reduction gearing, or speed reduction by use of a series of two trains of gears. Single reduction gearing is used in older and special installations where ship speed is fairly high. Gear teeth are of involute form, with the pinion teeth of harder material than the gear. The gear trains are of the double helical type to avoid heavy axial thrust (see Fig. 2). Double reduction gears of articulated type are constructed with flexible couplings between the high-speed train and the low-speed elements.

Another design, the nested gear, has the high-speed gear wheels outside the large low-speed bull gear or in between the separated helical halves of the bull gear.

Mechanical reduction gears are carefully constructed to close tolerances. They have forced lubrication in sprays ahead of the meshing teeth, to the bearings, and to the flexible couplings. Tests have shown that bearings represent at least half the power loss of the entire gear set.

Turboelectric drive. This type of drive, comprising one or more steam turbine generators and ac propulsion motors, is also used for ship propulsion in powers over 3000 shp. Ease of maneuvering has

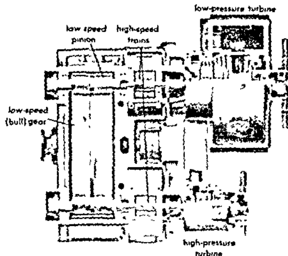


Fig 2. Double reduction articulated gear design for use with modern high-rpm marine steam turbines. (General Electric Co.)

favored the use of this drive for Coast Guard cutters and dredges. It was fitted to many United States tankers during World War II because of available manufacturing facilities. The synchronous motors are provided with induction winding for starting and reversing. Relatively large changes in propeller revolutions are made by alteration of the turbogenerator speed.

Motor, generator, exciter, and cooling equipment losses result in several per cent lower efficiency than is the case with geared steam turbines.

Diesel-electric drive. This type of drive, composed of one or more dc diesel generator sets and often a double armature propulsion motor, is used in tugs, dredges, Coast Guard cutters, and icebreakers, where maneuvering and a wide range in propeller speed are necessary. For slow-speed operating during maneuvering, the engine speed is often reduced and the generator field excitation altered to provide wide variation in the motor output and propeller speed.

Control arrangements. Bridge or pilot-house control of the ship propulsion unit, without action by the engineer on watch, is feasible with diesel-electric drive and for small, low-powered direct-drive and mechanically geared diesel installations. This is the customary arrangement for tugs and dredges. With steam propulsion and medium and large direct-drive diesel installations, the engineer carries out orders transmitted from the bridge and reports his action. He is in a position to check associated vital equipment and auxiliaries, such as the boiler plant, starting air pressures of the diesel, lubrication systems, pumps, and so on.

Governors. Above the operating rpm, ship propeller torque increases faster than engine or turbine torque, and thus ship propeller drive is inherently stable. Because of the ship's pitching, the propeller may lift partially out of the water a the engine may tend to race. To allow for this a . for propeller shafting failure, A

Bureau of Shipping regulations require that a governor be fitted to limit over-speed to 15% above the rated speed.

A common type of governor uses oil pressure developed by small pumps incorporated with the main turbine rotors to activate the governor; low lubricating oil pressure will also shut off the steam supply.

With turboelectric and diesel-electric drive, there is no mechanical connection between the generator set and the propulsion motor and propeller. The operating governor holds generator speed at the set value by throttling turbogenerator steam or the amount of fuel injected in the diesel engine cylinders [E.BU.]

Bibliography: T. Baumeister (ed.), *Marks' Mechanical Engineers' Handbook*, 6th ed., 1958; J. M. Labberton and L. S. Marks, *Marine Engineers' Handbook*, 1945, H. L. Seward (ed.), *Marine Engineering*, 2 vols., 1942–1944.

Marine engineering

Marine engineering embraces the design, construction, installation, operation, and maintenance of main power plants as well as the associated auxiliary machinery and equipment for the propulsion of ships. In general, most marine engineering practices have been proven ashore before being adapted to shipboard use.

The areas of interest of the marine engineer and the naval architect overlap in so many matters that a sharp line of distinction is not possible. However, the naval architect is principally concerned with the ship's hull, and the marine engineer with the machinery, particularly the main and auxiliary propulsion machinery and the associated systems and equipment.

Since a sea-going ship must operate basically as a self-sufficient and independent community, the range of engineering problems is very broad. Most important of all is the fact that the propulsion plant must always continue to propel the ship. Modern technology offers several types of power plants in various combinations, all the way from the old reciprocating steam engine to nuclear types. In addition, modern electronic, electrical, and refrigeration equipment is necessary to meet ship operational requirements as well as to afford customary living arrangements. As a result of this complex technological picture, marine engineering design has tended to break up into a number of specialties. In spite of complex problems faced by designers, they have continued to design more efficient plants which are capable of reliable operation by a minimum number of engineers. *See* MARINE BOILER; MARINE ENGINE; MARINE MACHINERY; PROPELLER, MARINE; REACTOR, SHIP PROPULSION; SHIP PROPULSION. *See also* NAVAL ARCHITECTURE. [K.K.C.]

Marine fisheries

The harvest of organisms of the sea to provide both food and organic materials for agriculture and industry. The harvest, amounting to 26,000,000 metric tons of fish and marine invertebrates in 1956, can be greatly increased to supply a substantial share of the protein food for the world's growing human population, since many of the living resources of the sea remain unutilized.

Production of sea fisheries for 1956*

Ocean zones	Metric tons, millions	%
Arctic	1.2	4 6
Temperate, Northern Hemisphere	18 6	70.7
Tropical	5.0	19.0
Temperate, Southern Hemisphere	1.5	5.7
Antarctic	0	0
	26.3	100 0

* Excluding whales.

For further information concerning fisheries techniques, management of fisheries, fishery products and processing methods *see* FISHERIES CONSERVATION; FOOD MANUFACTURING.

Production of sea fisheries. The recent growth of the world's fisheries is illustrated in Fig. 1, which shows the harvests of the major categories of marine organisms, and includes also, for comparison, the yields of the fresh-water fisheries. In addition, there is an annual whale catch of about 4,000,000 metric tons, mostly from the Antarctic, which substantially utilizes the full potential yield of the whale populations. A tabulation of the production, other than whales, by zones of the oceans is given in the accompanying table.

It is evident that the bulk of the catch is taken from waters of the Northern Hemisphere, despite the fact that the southern oceans constitute 57% of the world's sea area. This is related to three fac-

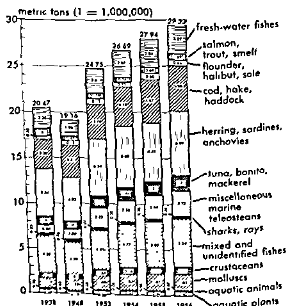


Fig. 1. The world catch of fishes and marine invertebrates. (From L. P. D. Gertenbach, ed., *Yearbook of fishery statistics, 1955–1956*, Food and Agriculture Organization of the United Nations, 1957)

tors: (1) human populations, and thus markets, are heavily concentrated in the Northern Hemisphere; (2) the major fishing nations are the industrialized maritime nations, which are located there; and (3) except for some of the tuna, salmon, and herring fisheries, the important fisheries are located in the relatively shallow sea areas over the continental shelves, the extent of which is much greater in the Northern than in the Southern Hemisphere. The second point may be illustrated by tabulating the catches, in thousands of metric tons, of the six leading nations in 1956, amounting to 61% of the world total:

Japan	4,763
United States	2,936
China (mainland)	2,640
USSR	2,620
Norway	2,129
United Kingdom	1,050

These are all nations of the Northern Hemisphere, and, except for a small part of the catch of tunas made by the United States and Japan, their catches come entirely from that hemisphere.

The locations of most of the established fishing grounds of the world are shown in Fig. 2, which indicates that most of the principal fisheries are located within a few hundred miles of land.

An additional noteworthy feature of the sea fisheries is that, in each area of the sea, the bulk of the harvest consists of a relatively few species, for which there is a good market, among the many present. In the United States, over 125 kinds of fish are listed in the 1956 catch of 2,100,000 metric tons (excluding crustaceans and mollusks), but a dozen of these accounted for 1,500,000 metric tons, or about 70%. These are (in thousands of metric tons):

Menhaden	943
Salmon (5 kinds)	123
Sea herring	116
Yellowfin tuna	69
Skipjack tuna	56
Haddock	69
Atlantic Ocean perch	69
Pacific halibut	19

Of these, it is probable that the salmon, haddock, yellowfin tuna, and Pacific halibut are being exploited at or near the maximum productive capacity of the stocks. For the herring, skipjack tuna, and ocean perch, on the other hand, it is quite certain that the harvests can be increased.

Similarly, in the United Kingdom 75% of the catch of marine fish consists of six demersal species (plaice, cod, haddock, hake, saithe, and whiting) and one pelagic species (herring). It appears that the stocks of the demersal species, except cod, are being nearly fully utilized. The pelagic species, however, both the herring and other kinds now little used, such as sprat, pilchard, and mackerel, could provide much larger catches.

In 1949 a conference of fishery experts meeting under the auspices of the United Nations concluded that in every part of the world there are known large fish stocks which are not being utilized appreciably. Other stocks will likely be discovered in the future, especially in the little-explored offshore waters of the high seas. Their exploitation requires only the economic incentive and technological development to make it profitable.

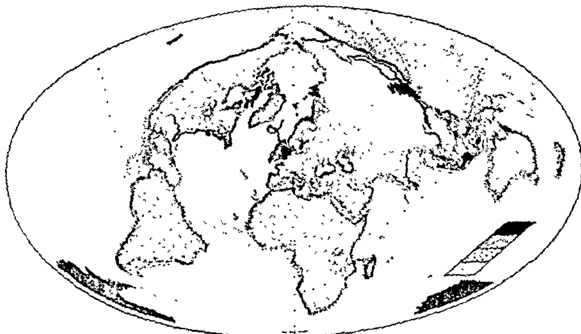


Fig 2. Distribution of fishing grounds of the world. The yield of the sea is suggested by the relative intensity of shading. (Adapted from L. A. Walford, Living Resources of the Sea, Ronald Press, 1958)

Potential harvest of sea fisheries. The present state of knowledge of living resources in the sea is inadequate for any precise estimate of the potential yield, but it may be many times the present harvest. However, it is undoubtedly less per unit area than the harvest of the land, because of the different conditions of life in the sea.

Comparison with land life cycle. The cycle of life on land is relatively simple; it may be described in four stages. First are the plants, which by the complex chemistry of photosynthesis use the energy of sunlight to build organic matter. Herbivores, such as sheep and cows, live on the plants; carnivores, such as tigers and men, eat herbivores. Finally the bacteria decompose the dead bodies and excreta of all living creatures, making their constituent substances again available as building materials for the plants. In the sea, the cycle is much longer. The plants in the open sea are microscopic phytoplankton; in place of cows, the zooplankton animals that eat the plants are tiny organisms, such as Crustacea, some no bigger than the head of a pin. Many kinds of carnivores eat the herbivores, but they are mostly zooplankton too, only a fraction of an inch long. Other intermediate flesheaters exist between them and the fishes of our ocean harvest. Because each link in this long food chain is inefficient, man can reap from the sea only a small fraction of its organic production.

The harvest from the sea is also limited, as compared with that from land, by other factors. Sunlight which falls on the sea can be used by the plants throughout the upper 20–100 meters, thus providing a large living space for the plants. The great areal extent and volume of the ocean, and its fluidity, result in low [illegible]

per ur[illegible]

Furthe[illegible]

and animals are of the same order of magnitude as the amount of the annual organic production, whereas in the ocean the standing crop is small compared to the production, because of more rapid turnover. Average rate of organic production per unit area is probably about the same on land and sea, but efficiency of harvesting depends more on the size of the standing crop than on the total production of organic matter.

Role of plants. All life in the sea, as on the land, is supported by plants, which grow only in the waters near the surface where bright sunlight penetrates. These near-surface waters differ widely in fertility; the ocean has its green pastures where life flourishes in abundance, and its deserts where a few poor plants and animals barely survive. Fertility of the land depends on temperature, sunlight, water, and plant nutrients. In the sea, water is always abundant; the plants are well adapted to the narrow range of annual temperature variations; the intensity of sunlight determines the length of the growing season, and the depths in which plants grow, but usually not differences in fertility. These differences depend primarily on the plant nutrients in the sunlit waters near the surface. The nutrients are always abundant in the deeper layers of the sea, where they are accumulated by the slow sinking and decomposition of the plant and animal remains from the brightly lighted near-surface layer. The fertility of the upper layer is restored when the deep, nutrient-rich waters are brought near the surface. Those regions where the deep water is most effectively plowed back into the surface water are the green pastures of the sea. *See* SEA WATER FERTILITY.

The plowing is accomplished in three ways. In some regions the wind drives the surface waters away from the coast or away from an internal boundary, and in such regions nutrient-rich waters are upwelled from mid-depths. In high latitudes, surface waters are cooled in the winter and, becoming more dense, sink and mix with the deep water. In other places, violent mixing along current boundaries brings deeper waters into the sunlit zone. *See* OCEAN CURRENTS; SEA WATER.

Most of the world's great oceanic fisheries are located where basic organic production, fostered by one or more of these processes, is high. In similar, yet-unexplored areas of high fertility, new, rich fisheries can certainly be found. One such example occurred in 1952–1957: the development of a new high seas fishery for tuna along the African coast and across the equatorial Atlantic, associated with regions of upwelling. Thus, the mapping of the ocean circulation and associated variations in basic fertility may provide guides to the efficient discovery of additional rich fishing grounds.

Improving the fishery harvest. Fishermen will catch fish only if it pays to do so, that is, if the value of the product is at least equal to the cost of production. The fuller exploitation of unutilized and underutilized stocks of fish and marine invertebrates depends, therefore, on decreasing the cost of production, increasing the value of the product, or both.

Decrease in cost of production can come through the discovery of previously unknown abundant stocks of commercially valuable species, such as the recent discoveries of the king crab of the Bering Sea or the tuna stocks of the tropical Atlantic. There is also need, however, for improved harvesting techniques. Although modern fishing craft use the most advanced aids to navigation, and some use has been made of underwater sound equipment and aircraft for locating fish schools, the various types of nets, lines, and trawls used for harvesting them are basically the same as have been employed for centuries. Consequently, the major fisheries are for those few species whose aggregation habits make them amenable to capture by traditional kinds of gear. Other fish species, occurring in the sea more diffusely than those now exploited, may well constitute a much larger biomass and be capable of yielding greater harvests. If these are to be commercially harvested, it will be necessary to develop new means of capture, possibly, for example, the

use of electrical or acoustical fields to cause them to aggregate so that they can be netted. Unfortunately, the fundamental knowledge of the behavior and reactions to stimuli of these organisms, upon which such technological developments can be based, is inadequate. Even for the presently exploited fish stocks there appears to be much need for the development of new and more efficient harvesting methods.

Increasing the value of the products of the fisheries depends, fundamentally, on increasing the demand for them among consumers. For the large share of the harvest that is used for human food (about 85% in 1956) this requires better processing, distribution, and marketing, and creation of new food products of greater consumer acceptance. It also includes further development of uses other than as human food. A considerable share of the fish catch is now used for making fish meal and oil, the former being used mostly for feedstuffs for chickens, cows, and other domestic animals, while the oil is used in a variety of industrial products. The share of the world harvest used to make meal and oil has increased from 8% to 14% between 1948 and 1956.

Conservation problems. The populations of fishes and other organisms supporting the marine fisheries are self-renewing resources, and thus are capable of being exploited by man in perpetuity. To preserve such resources it is not necessary to refrain from using them, but only to use them in such fashion that they remain capable of yielding maximum benefits in the future. Since a fish stock is not only self-renewing, but the rate of renewal is dependent on the share left to perpetuate itself, which is reduced by the harvesting, it has the peculiar property that the sustainable harvest increases with increasing fishing effort up to a certain point, when the maximum average catch is obtained, and then actually decreases with increasing fishing effort. This is shown graphically in Fig. 3, which shows schematically the relationships among fishing intensity, average stock size, and average total annual catch.

Since the stock size decreases with increased fishing intensity, and thus the returns per unit of effort (and cost) decrease, it is almost never economically possible to drive a marine fish population to extinction, since it becomes uneconomic to fish long before this. The conservation problem is not, therefore, to prevent the stock from being exterminated, but to maintain it at a desirable level. In general, it is accepted that the level of maximum sustainable total catch is the optimum, and that overfishing occurs only when the intensity of fishing exceeds the corresponding value.

Most of the populations of organisms supporting marine fisheries are underfished. However, the intensity of fishing for some of the most desirable kinds of fish and other marine organisms is at or beyond the level of maximum sustainable yield, and for others this point is likely to be reached in the near future. This is true, for example, of several of the demersal fish stocks of the North Atlantic, most whale stocks, the halibut and salmon of the North Pacific, and the yellowfin tuna of the eastern tropical Pacific. A great deal of the effort of fishery scientists of governmental agencies, and of such international organizations as the International Council for the Exploration of the Sea, the International Halibut Fisheries Commission, the Inter-American Tropical Tuna Commission, and the International Whaling Commission is devoted to ascertaining the degree of exploitation of the stocks supporting important marine fisheries, and to devising measures for preventing overfishing where it is imminent, or for effectively curtailing the fishery where it has already occurred. This requires extensive studies of the production statistics of the fisheries and of the vital statistics of the fish populations, in order to estimate the effects of fishing on the populations, and requires studies of the many complex details of their ecology upon which to base efficient conservation measures.

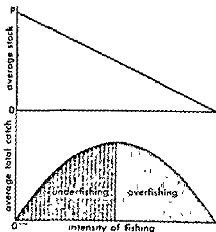


Fig. 3. Relations among fishing intensity, stock size, and average sustainable total catch from a fish population.

It is, as noted above, generally accepted that the objective of conservation is to maintain the fish populations at such levels that they are capable of providing the maximum sustainable total catch. It has been argued by a number of economists, however, that the maximum economic yield, that is, the maximum of the difference between the cost of taking the harvest and its value, is a more reasonable objective for fishery management. If so, the attainment of such an objective requires taking into account not only the biology, ecology, and population dynamics of the fish populations but also the economics of commercial sea fishing. The latter is rather more difficult than the economics of utilization of terrestial natural resources, because the fish populations mostly lie on the high seas, where they are accessible to fishermen of all nations on an equal basis; even where they are confined to the territorial waters of a single nation, they are usu-

ally treated as the common property of all its citizens. *See* MARINE RESOURCES. [M.B.S.]

Bibliography: L. P D. Gertenbach (ed.), *Yearbook of Fishery Statistics, 1955–56,* 1957; M. B. Schaefer, *The Scientific Basis for a Conservation Programme*, in Papers of the International Conference on the Living Resources of the Sea (U.N.), 1956; R. Turvey and J. Wisemen (eds.), *The Economics of Fisheries,* 1957; United Nations, *Proceedings of the United Nations Scientific Conference on the Conservation and Utilization of Resources,* vol. 7, 1951; L. A. Walford, *Living Resources of the Sea,* 1958.

Marine influence on weather and climate

The moderating, energizing, and redistributing effects that bodies of water have on air temperature, weather, and climate by releasing energy and heat to the atmosphere in characteristic ways.

To a greater extent than is commonly realized, the climates of the oceans determine the climates of the earth. In meteorological literature, however, discussions are mostly limited to a few words on the moderating effects of oceans on the climate of continents It is often overlooked that oceans constitute the prime source of energy for all water-vapor-induced atmospheric processes. The latent heat liberated by the condensation of water vapor evaporated from the oceans constitutes the prime source of energy for cyclone formation; and such water vapor from the oceans is otherwise a primary stage in the energy cycle of the general atmospheric circulation *See* STORM.

When the climatologist speaks of the marine influence on climate he is usually referring to the moderating effect of the oceans rather than to *their role in the energy cycle of the* atmospheric circulation. The usual explanation given for the moderating effect of the *oceans is that sea* water has a higher specific heat than does the soil surface *of the continents*. This factor, however, is only one *among a number* of others of equal or greater importance that serve to impose a lag in the heating of the surface layers of the *atmosphere.*

These factors include: (1) *The largest fraction* of the solar energy absorbed at *the sea surface* is utilized in evaporating sea water; a very much smaller fraction (average about 10%) is utilized in directly heating the atmosphere in *contact with* the sea surface. (2) A large *proportion* of the surplus solar energy absorbed at the *sea surface during* summer at mid latitudes is stored by the oceans, to be released to the atmosphere *during the colder* seasons when there is a deficiency *in solar* energy. By the same mechanism much of the solar *energy* absorbed at low latitudes is transported by ocean currents and released to *the atmosphere* at higher latitudes where receipt of *solar energy* is also comparatively deficient. (3) *The energy absorbed* at the sea surface is mixed *throughout* a relatively deep layer whereas the energy *absorbed at a* soil surface is confined to heating a shallow *surface* layer. Neither superheated nor subcooled surface zones remain unaltered for any length of time over the open ocean. (4) Shorter wavelengths of energy which reach the sea surface are transmitted to considerable depths because of the relative transparency of sea water to such radiation; but this is of secondary importance in the energy budget of the oceans. (5) Less solar energy is available for absorption over the sea than over land because of the higher albedo of the ocean areas. The high albedo results primarily from the greater average cloudiness rather than from a high reflectivity of the sea surface (*see* ALBEDO).

The oceans affect weather and climate, then, because they are a tremendous reservoir for the storage of solar energy. This energy is absorbed during periods of the year (or in regions) of surplus solar energy. It is then stored or transported by ocean currents, to be released to the atmosphere at times (or in regions) of deficiency in solar energy. In this sense the oceans become equalizers or moderators of climates.

See CONTINENTALITY, WEATHER AND CLIMATE. [W.C.J.]

Bibliography: W. C. Jacobs, On the energy exchange between sea and atmosphere, *J. Marine Research,* 1942; G P. Kuiper (ed.), *The Solar System,* vol. 2, 1954.

Marine machinery

Practically all the machinery used on ships has shore-based counterparts. It is often developed for marine use prior to shore development and use. There are many special demands placed on marine machinery. These include an extra measure of reliability, a rolling and pitching platform, limited space and weight considerations, as well as conditions favorable to corrosion. These factors make necessary the *frequent use of special* designs for marine service. In addition, the variety of purposes served by different types of ships and boats call for a number of types and characteristics of machinery *for best results.*

This article discusses general characteristics of ship propulsion machinery, types of propulsion, auxiliary machinery, shipboard electric plants, the arrangement of machinery items, and materials used in machinery. For additional information, *see* MARINE BOILER; MARINE ENGINE; PROPELLER, MARINE; REACTOR, SHIP PROPULSION; SHIP PROPULSION.

PROPULSION MACHINERY

Marine propulsion engines, except for those used in smaller craft, are basically either of the steam or the diesel type. Each type may be directly connected to the propeller shaft or connected through some type of reduction gear. Many smaller craft use gasoline engines.

Since high efficiency of propellers is not compatible with their operation at high rpm, some type of speed reduction between the prime mover and the propeller is commonly used. Older steam plants, as well as some still being installed, use

Scotch or fire-tube boilers with reciprocating steam engines which are directly connected. However, for more efficient operation of both the propellers and the power plant, later type steam installations use water-tube boilers and steam turbines with mechanical reduction gears or with electric drive for accomplishing the speed reduction. Diesel engines are also used with direct connection to the propeller, but a mechanical reduction gear or electric drive is usually employed for the reduction function.

Gas turbines and free-piston engines have been used in experimental marine plants. In marine gas turbines the fuel is injected and mixed with air from a compressor, after which it is burned. The resulting hot gases are allowed to expand in a turbine which drives the ship and the compressor With free-piston engines the combustion system, or gasifier, and the compressor are combined in several free-piston units which are essentially 2-cycle opposed-piston diesel engines.

Nuclear power plants have also pioneered in marine applications, especially submarines In these power plants, the conventional steam boiler and the fuel oil supply for propulsion are replaced by the reactor and the nuclear fuel, respectively. The remainder of the plant is generally similar to a conventional steam plant.

Reliability of machinery. Marine machinery must be reliable in operation; otherwise the services of the ship, or even the ship itself, may be lost Other considerations such as fuel economy, initial cost, space, and weight are important but they are secondary to reliability. This fact applies not only to the main propulsion machinery but to much of the auxiliary machinery and equipment. Duplicate equipment is often installed as a safety measure. As an alternative, emergency means of accomplishing required results may be provided. However, a failure of a necessary operating component at a critical time can still lead to serious difficulties.

Fuel considerations. Fuel oil is much more generally used than coal as the fuel for marine power plants, though coal still finds use on some short runs where it is readily available, such as on the Great Lakes. Fuel oil occupies much less space for the same capacity and is much cleaner and more conveniently handled and stowed than coal The cost of any type of fuel varies with supply and demand, location, and other factors. For example, cost is normally higher in remote ports and the cost of diesel fuel oil may be 50–100% higher than fuel oil suitable for use in a steam plant. The rate of consumption of fuel oil varies considerably with different types of propulsion plants, while even among plants of the same type there is appreciable variation in fuel oil consumption. Figure 1 shows average rates of fuel consumption for several types of marine propulsion plants. Steam turbine plants using higher temperatures and pressures will give better results than those indicated in Figure 1. Fuel consumption also varies with the percentage of normal power being developed. For example,

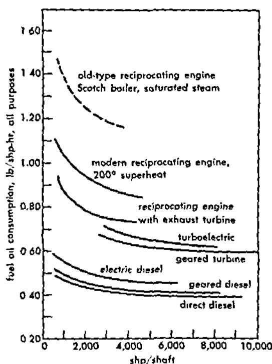


Fig 1. Fuel consumption for various propulsion plants. (*Society of Naval Architects and Marine Engineers*)

since merchant ships normally operate at nearly full power their power plants are designed for best economy at this speed. In the case of naval vessels, however, since their normal cruising is done at comparatively low speeds, their plants are designed for best economy at lower speeds.

Weights and costs. Since the weight of machinery for ships runs from about 15% of the total light weight (light weight is the weight of the vessel itself) for cargo ships to about 50% for naval vessels, the importance of weight is obvious. For merchant vessels the relative weight of machinery for different types of power plants is indicated in Fig. 2. For naval vessels of higher power, where weight saving is of great importance, the weight of machinery is about 25–50 lb per shaft horsepower (lb/shp). All ship's machinery is included in Fig. 2, except deck machinery, steering gear, windlass, and piping not directly associated with the propelling machinery and its auxiliaries.

Figure 3 indicates average trends of relative costs in machinery. Actual costs vary with costs of labor and material and other factors.

Vibration and noise. Mechanical unbalance of either the propellers or any of the components of the operating (running) machinery will cause vibration. Another source of vibration is the variation in forces transferred to the hull by the rotating blade's pressure field. Torque reactions from reciprocating engines, either steam or diesel, may also produce vibration. Twin-screw ships are subject to vibration from slight variations in speed between the engines or propellers which result in beat synchronization (*see* BEAT).

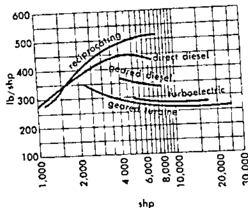


Fig. 2. Weight of machinery for various power plants. (*Society of Naval Architects and Marine Engineers*)

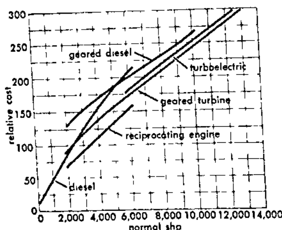


Fig. 3. Relative costs of machinery. (*Society of Naval Architects and Marine Engineers*)

Vibrations caused by unbalance are eliminated by balancing the faulty item. This is more easily accomplished with rotating than with reciprocating machinery, though special equipment and techniques are required in either case. Proper clearance between propeller blade tips and the hull is helpful in eliminating vibration from this source. Vibrations which result from torque reactions in reciprocating engines are sometimes eliminated by making suitable brace connections between the engines and the ship's hull structure or by installing struts or braces between two engines. Flywheels and counterweights are often useful on steam and diesel reciprocating engines. Speed-governing devices are useful in eliminating vibrations from beat synchronization.

High-speed rotating machinery components, especially mechanical gears, often produce vibrations in the audible frequency range. These vibrations, if they are of sufficiently high amplitude, may produce objectionable noise. In many cases where the effects of such noise are unpleasant and elimination at the source is not possible, acoustical treatments are available. Soundproofing involves special construction features which reduce the transmission of noise. Sound-absorbing materials are also effective when properly installed. *See* Noise control.

Maintenance. Since a ship must operate as a self-sufficient and independent community, the proper maintenance of machinery takes on special importance. In order to recognize and correct irregularities before they develop into troubles, it is common practice to establish a regular maintenance program. Fixed routine checks, tests, and procedures are desirable for both in-port and at-sea conditions. Repair parts may present a problem, which should be kept in mind as a factor in choosing the type of machinery for use on a particular ship.

Diesel propulsion. Figure 1 indicates that the diesel engine has the lowest fuel consumption rate of any of the several types of marine power plants. Figures 2 and 3 indicate that, up to about 2000 hp, the cost and weight of diesel engines compares favorably with other types. Among other advantages of diesel engines are quick starting, stopping, and securing, as well as reduced standby losses. Disadvantages include higher maintenance costs due to the high pressures and temperatures encountered in the cylinders; these create lubrication problems and a correspondingly high consumption of lubricating oil.

When the diesel engine is directly connected to the propeller, the slow speed required for best propeller efficiency makes for a heavy engine with large pistons and other parts. For example, Fig. 4

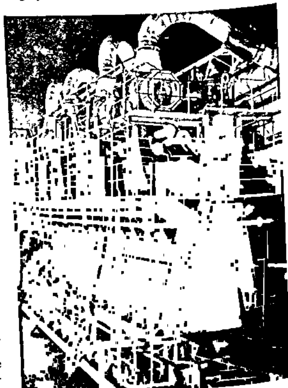


Fig. 4. Modern 15,000-hp direct-connected marine diesel. (*Burmeister and Wain*)

shows a very large diesel engine built in Europe by Burmeister and Wain in 1958 for installation in a 32,000-ton tanker. This engine has 12 cylinders and is supercharged to develop 15,000 brake horsepower (bhp) at 115 rpm, a much greater power than the maximum indicated by Figs. 1, 2, and 3.

In the United States, direct-drive diesels are used to some extent, but speed reduction for propulsion is a more common arrangement. For diesel propulsion a mechanical reduction gear with a speed ratio of 2 or 3 to 1 between the engines and the propeller is often used. For such an installation some form of flexible coupling is installed between the engines and the gear pinions in order to absorb the variations in torque, thus preventing vibrations. The reduction in speed makes for lighter, higher-speed engines which are easier to overhaul and which offer reduced vulnerability in cases where more than one engine is used. Figure 5 shows an 8-cylinder 4-cycle marine diesel of 1420 hp at 458 rpm with a pneumatic flexible coupling and a 2.54:1 reduction gear, both at the right end.

Speed reduction for diesel or steam propulsion may also be accomplished by means of electric drive. In diesel-electric propulsion, two or more diesel-driven generators usually supply direct current to the main propulsion motor. The engines run continuously in one direction at constant speed under governor control. Reversing is accomplished electrically and speed of the propeller is controlled by varying the field strengths of the generators and main motor. Excellent maneuvering qualities are obtained, along with elimination of starting, stopping, and reversing of the engines during maneuvering. In addition, the location of the diesel generators is a matter of choice and entirely independent of the location of the propelling motor. Also, this type of drive is easily adapted to bridge control, which is desirable for maneuvering in close quarters. Thus the diesel-electric drive, though relatively high in initial cost, makes an excellent propulsion plant for tugs, ferries, dredges, and other craft where good maneuverability is necessary.

Steam propulsion. The original steam engine, as developed by Thomas Newcomen for pumping water out of mines in England, was improved by James Watt in the 1760s and was adapted to ship propulsion by Robert Fulton in 1807. Steam reciprocating engines embodying all the features developed by Watt (plus some refinements) are still in use. The boilers originally used for generation of the steam were generally of the Scotch or fire-tube type, burning coal. Later boilers are of the express or water-tube type, burning fuel oil, though Scotch boilers are still used in some cases and may burn either coal or fuel oil. The Scotch boiler, like the reciprocating steam engine, is simple and rugged, and the long continued use of the combination demonstrates the confidence of the marine industry in the reliability of this type plant.

The physical size of the larger reciprocating steam engines developed in the early 1900s was impressive Some engines had a low-pressure cylinder diameter of 112 in. and a stroke of 6 ft. Reciprocating parts of such size introduced many problems. With a continuing demand for higher powers, more attention was given to the development of steam turbines, which are inherently better adapted to the use of higher vacuum, higher temperature, and higher pressure steam.

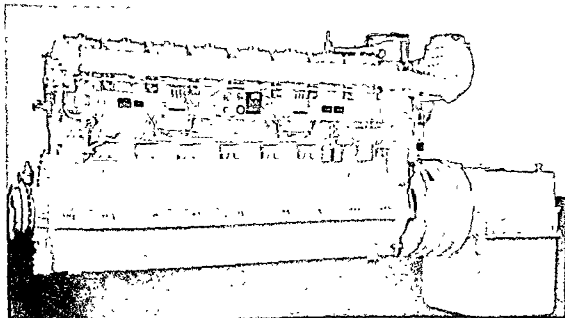


Fig. 5. Marine diesel of 1420 hp with flexible coupling and reduction gear at right end. (Copper-Bessemer Corp.)

Steam turbines, as originally patented by C. Parsons in England in 1884 and G. de Laval in Sweden in 1888, and as further developed in the 1890s by C. E. A. Rateau in France and C. G. Curtis in the United States, passed through various stages of development. Turbines, including features from all four early designers, became the most popular steam propulsion plants. However, since they inherently operate at high speed for best efficiency, the use of some form of speed reduction between the turbine and the propeller is a necessity. As in the case of the diesel engine, mechanical reduction gears are generally used and speed reduction ratios of 20:1 are common with a single-gear train and 80 or more to 1 with double reduction gears. For reversal of engines, a separate astern turbine is usually installed in the exhaust trunk of the ahead low-pressure turbine, as shown in Fig. 6, which pictures a typical modern double reduction gear unit with high- and low-pressure turbines, the top half of the turbine and gear case having been removed. The astern turbine may be seen at the far end of the low-pressure turbine. The steam, for going ahead, undergoes partial expansion in the small high-pressure turbine to the right and then completes expansion in the larger low-pressure turbine to the left in the figure. The torque developed by the astern turbine is approximately 80% of ahead torque. Turbines, like diesel engines, are also well-adapted for use with electric drive.

Nuclear propulsion. The United States Navy pioneered the development of nuclear power, primarily for submarines. An experimental shore-based prototype built at Arco, Idaho, was the first practical nuclear power plant of appreciable power output. The first such plant to be used in any ship was installed in the United States Navy submarine *Nautilus*, which commenced operation in January, 1955. With a nuclear plant, air for combustion is no longer necessary; hence very long submerged endurance became an immediate great advantage to submarines. At present, nuclear power costs somewhat more than conventional power because of expensive new materials and special techniques. This

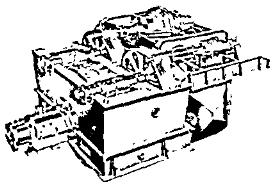


Fig. 6 Modern marine steam turbine propulsion unit with double reduction gear. The top half of the turbine and gear case has been removed. (Westinghouse Electric Corp.)

cost disadvantage is readily overcome by the gain in military performance in naval ships. Submarines have cruised submerged for months, whereas with other plants only several hours of continuous submergence is possible. A nuclear power plant of 22,000 shp capacity is installed in the American ship NS *Savannah*, the first merchant vessel to be so equipped.

Shafting. A typical arrangement of shafting as installed on a twin-screw high-speed naval ship is shown in Fig. 7, with principal parts identified. In the case of twin-screw merchant ships the portions of the shaft which would be external to the hull are sometimes covered by the shell-plating of the hull, the plating being bossed out around these external shaft sections. This arrangement offers protection from mechanical damage and exposes less shafting to the corrosive action of sea water. In addition, the size of propeller and stern tube shafts is made somewhat larger than the line shaft in order to afford a margin of strength in a vulnerable and inaccessible area, to permit some corrosion with safety, and to withstand better the higher loads on these sections. Serious flexural and torsional vibrations may occur in shafting; however, the size and characteristics of the shafts and connected rotating system determine the values of the critical flexural and torsional speeds. The design, made to meet classification society rules, is such that the flexural critical speed is greater than the maximum operating rpm and also such that no severe torsional vibrations occur within the operating range of speeds.

Determination of horsepower. The power required in a vessel depends on its resistance to propulsion at its operational speeds. A rough approximation of the power required is first made, based on the general dimensions of the vessel, in order that weights and spaces may be assigned for machinery and fuel. As the design proceeds, the approximations are refined until final values based on the final dimensions and form of the vessel are established. *See* SHIP DESIGN.

Several methods of approximating and finally estimating shaft horsepower are used, including the Admiralty coefficient, the © coefficient, standard series, and model testing. Values of coefficients are affected by many things, such as the proportions of the ship, the speed-length ratio, and the propulsive coefficient. Extensive records of actual performance of other vessels or models are required in selecting a suitable coefficient value for the vessel under consideration. D. W. Taylor's standard series is in common use by naval architects in arriving at shaft horsepower. It consists of published results of extensive resistance measurements of a series of models representing ships of a particular form with a wide range of proportions. Towed and self-propelled model tests are useful in accurately determining horsepower at various corresponding speeds; however, before the models can be made, the form of the underwater body and a definite line drawing of the ship must first be available.

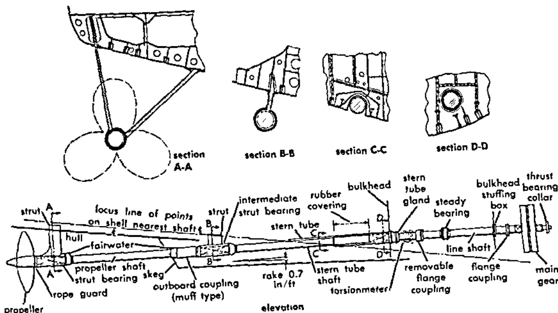


Fig. 7. Twin-screw shafting arrangement. (Society of Naval Architects and Marine Engineers)

The estimation of horsepower requirements by the various methods indicated applies in ideal conditions such as smooth seas and clean hulls. The service conditions encountered involve semifouled bottoms and actual sea conditions which require 20–35% more power than the estimates.

AUXILIARY MACHINERY

The operation of the main propulsion plant requires the use of various auxiliaries, depending on the type of plant. Other auxiliaries, not connected with the operation of the main plant, are used to provide hotel services (heating, air conditioning, wiring, hot and cold water piping, sanitary drains) or other necessary functions.

Heat exchangers. All types of marine power plants utilize heat exchangers of various types and for various purposes. In steam plants the largest heat exchanger is the main steam condenser, whose purpose is to serve as an evacuated receiver for liquefying the spent steam exhausted by the main engines, thus conserving the condensate for return to the boilers where it is changed to steam again. The large size of the main steam condenser ... vol- high

Internal construction features of condensers vary considerably in design. A typical marine surface condenser is shown in Fig 8. The exhaust steam enters through the large opening at the top and passes down over a large nest of uniform-diameter (⅝- to 1-in.) tubes, through which sea water is flowing. The sea water is carefully separated from

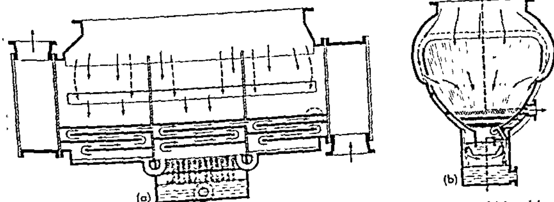


Fig. 8. (a) Condenser with secondary steam admission from bypasses, contracting-pitch tube spacing, baffled air cooler, and shrouded lane condensate reheating. *The sea water enters at lower right of (a) and leaves at upper left.* (b) Transverse view. (Society of Naval Architects and Marine Engineers)

the steam side. The path of the sea water is shown in Fig. 8*a*; the water enters at the lower right, passes through the tube nest, and leaves through the outlet at upper left. As the steam contacts the cooler tubes it condenses, the condensate finding its way to the lower chamber and then to the pump suction. Since any lowering of the condensate temperature below that necessary to produce condensation (undercooling) represents an overboard heat loss, design features are incorporated to reduce this. For example, in Fig. 8*b* it will be noted that baffles or shrouds are installed to direct some of the steam flow to the lower parts of the tube nest where it will contact condensate that has been undercooled, thereby raising its temperature (reheating) as the steam condenses. Air or noncondensable vapor is cooled, and its volume thus reduced, by the lower rows of tubes, the path being controlled by baffles to the air-pump suction which is located at the sea-water entrance, or coolest end. The separation of the sea water from the condensate is made difficult by the large number of tubes, which are places for possible leaks, as well as by the difference in pressure between the sea-water and condensate sides. However, this separation is very important since uncontaminated condensate has a high purity while there are many impurities in sea water. Boilers, turbine blading, and other components of a steam plant are sensitive to impurities in the water or steam and may be greatly damaged if such impurities are present.

Other types of heat exchangers found on ships include feed-water heaters, fuel oil heaters, lubricating oil coolers, and others. Many of these are of the shell and tube type shown in Fig. 9. The two fluids between which exchange of heat is to occur are separated completely from each other. One passes through the tubes while the other enters the shell and passes around the tubes. Baffles are often used, making the route devious or turbulent to improve heat transmission. In some cases tubes are replaced by coils of spiral shape. Parallel- or counter-flow of the fluids may be used, the latter being employed when the greatest change in temperature is desired. The higher pressure fluid is generally inside the tubes, though other factors, such as the corrosive qualities of the fluids and the relative amounts or volumes of each, are considered in selecting which is inside or outside the tubes. See HEAT EXCHANGER; HEAT TRANSFER.

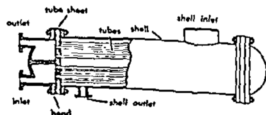


Fig. 9. Shell and tube heat exchanger. (Society of Naval Architects and Marine Engineers)

Pumps. Centrifugal pumps have many uses, and are available in a wide range of types and sizes. For example, the number of such pumps on a typical 10,000-ton passenger-cargo steam-driven vessel with an 8,500-shp plant may total 20–25, with capacities varying from 5 to 13,000 gal/min. The laws of similarity apply to geometrically similar centrifugal pumps, and permit the prediction of performance of a pump of different size but geometrically similar to one that has been tested. See SIMILITUDE.

Reciprocating steam pumps are generally double-acting; that is, they discharge from both ends of the pumping cylinder. The size of such pumps is set forth by giving first the steam cylinder diameter, then the pumping cylinder diameter, and finally, the stroke, all in inches. Simplex pumps have single steam and pumping cylinders connected by one piston rod. When two simplex pumps of the same size are arranged side by side, so that the valve motion of one is obtained from the piston rod of the other, the combination is called a duplex pump.

The term power pump is used to describe various pumps with reciprocating plungers or pistons of fixed or variable stroke that are driven by an external source of power, usually an electric motor.

Rotary pumps derive their positive-displacement-type pumping action from the rotary motion of the pumping element. They utilize different types of pumping mechanisms including gears, screws, cams and pistons, lobes, and vanes. Rotary pumps are used for higher viscosity fluids at moderate pressures and smaller volumes of flow.

Blowers and fans. In marine usage the term blower usually applies to induced- or forced-draft blowers for boilers, and the term fan refers to ventilation equipment, even though both devices may be similar in appearance. As in the case of centrifugal pumps, the laws of similarity apply to centrifugal blowers and fans. Propeller-type fans having two or more radial blades are also common.

Compressors and ejectors. Air compressors on merchant ships which have steam propulsion supply air at pressures of 100–150 psi for various purposes, including pneumatic tools, boiler soot blowers, air-operated control systems, and other general services. Figure 10 shows a commonly used compressor of the motor-driven type. For ships with diesel propulsion, high-pressure compressed air is required for engine starting and other purposes. The compressors are generally of the multistage reciprocating type with intercooling between stages. For scavenging air-supply, a rotary air compressor is often used.

Steam-jet air ejectors are used for removing air and noncondensable vapors from main and auxiliary steam condensers and, in some cases, from other types of equipment. Figure 11 indicates the principle of operation of the steam jet. When installed in air ejectors, two jets, or stages, are generally used, with an intercondenser between stages

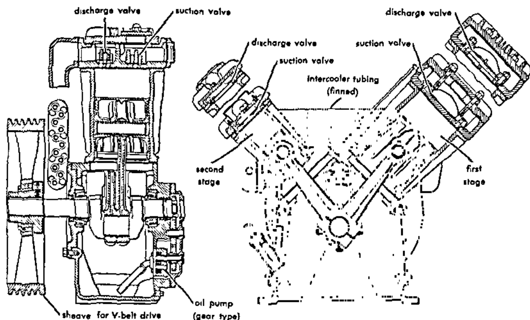


Fig. 10. Air-cooled type two-stage air compressor for ship's service compressed-air systems. (*Society of Naval Architects and Marine Engineers*)

to condense the steam used by the first stage, and with an aftercondenser for the second stage.

Distilling plants. These plants are used to produce pure distilled water from sea water. A common type of single-effect distilling plant as used on merchant vessels is shown in Fig. 12. The cylindrical evaporator shell contains an assembly of vertical steam-heating coils which can be removed as a whole for ease in cleaning or coil replacement. Steam is supplied at about 90 psig pressure, and the operating pressure of the vapor formed within the evaporator shell from the sea water being evaporated is about 5–10 psig. The cylindrical distiller condenses the vapor as it passes through removable vertical coils which are surrounded by circulating sea water. The capacity or size of such a plant varies with the size of the ship and the demands for distilled water. Where demands are large, more efficient multiple-effect plants are used. They operate with lower pressure steam, and the vapor from the first-effect evaporator is used for heating the coils of the second-effect evaporator, and so on, the vapor from the last effect only being condensed in the distiller. [K.K.C.]

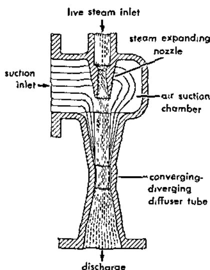


Fig. 11. Steam-jet air ejector. (*Society of Naval Architects and Marine Engineers*)

Ventilation machinery. Shipboard ventilation is necessary both for comfort and for cargo preservation. Ventilation is provided mainly by mechanical supply, exhaust systems, or both. When convenient, natural ventilation is used by means of weather deck cowls.

In machinery areas, watch stations are cooled by directional supply terminals, with a space time-of-change of about four minutes. Mechanical exhaust fans may be provided for hot spots.

Living quarters are ventilated by motor-driven centrifugal supply fans (direct-connected or belt-driven) usually located in fan rooms in the areas served. Exhaust fans serve toilets, galleys, and so on. Axial-flow fans are increasing in use. Heating is usually supplied by steam heaters in the supply ducts.

Air-conditioning equipment. Modern passenger liners are completely air-conditioned outside ma-

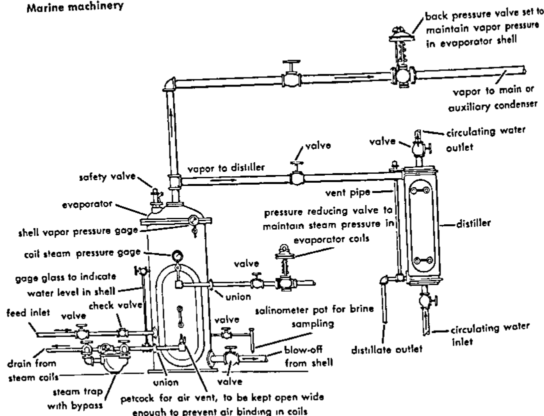


Fig 12. Single-effect distilling plant arrangement. (*Society of Naval Architects and Marine Engineers*)

chinery spaces, as are larger naval ships and the newer tankers and cargo vessels. Machinery used for air conditioning consists of a refrigerant compressor (Freon is generally used), a condenser (with salt-water cooling circuit and pump), and a cooling coil. Ships having isolated or few air-handling systems use direct expansion of the refrigerant in cooling coils. Where numerous systems are required, chilled-water heat exchangers are used. A pump circulates chilled water to the several cooling coils. Adequately dehumidified air is often too cold for comfort. For this reason air reheaters, controlled from the compartment served, are provided.

Compressors are motor-driven, and are of the centrifugal or reciprocating type, the former for capacities over about 125 tons. A passenger liner or aircraft carrier may have six or more 250-ton centrifugals, located in auxiliary machinery spaces. For a tanker, two 40-ton reciprocating compressors aft and one 30-ton reciprocating compressor midship is an average installation. Water chiller units have capacity control by means of automatic suction valve and hot gas bypass, and circuits are cross-connected between zones. Control of centrifugal compressors is accomplished by cylinder unloading and, for multiple units, by cross-connecting compressors.

Air revitalization equipment. Air filters of activated carbon provide the simplest means for odor control. However, as on submarines, air revitalizing systems also may include oxygen replenishment from storage bottles and toxic gas removal by carbon dioxide absorbers and carbon monoxide and hydrogen burners. [C.B.J.]

Bibliography: G. B. Johnson and D. A. Phillips, Air conditioning of modern tankers and cargo ships, *Trans. Soc. Naval Architects Marine Engrs.*, 494–517, 1956; Man-Made Air—The State of the Art, *U.S. Naval Inst. Proc.*, 30–37, 1959.

Steering machinery. The steering gear must meet the rudder torque requirements as determined by calculation or model test. For seagoing vessels, it must be powerful enough to rotate the rudder 35° to either side in about 15 sec at full speed ahead and in 30 sec at full speed astern.

Small vessels (tugs, large motor yachts, ferries, and so on) with power steering usually employ the cable and quadrant type of gear, either steam- or motor-driven.

Hydraulic drive, due to its superior flexibility and its adaptability in supplying stand-by power in an emergency, is fitted on most large, modern vessels, both naval and merchant. A variable, reversible-stroke pump, driven by a constant-speed electric motor, supplies oil under pressure to rams (pistons) which actuate the rudder crosshead.

With the exception of some small river and harbor craft, where the rudder moves in response to a pushbutton or lever, the rudder is controlled by a steering wheel. A telemotor, which is usually electrically operated, connects the steering wheel to the steering engine. A follow-up mechanism stops the steering engine when the rudder has moved an amount corresponding to the movement

of the steering wheel. The electric telemotor affords easy operation of the steering wheel and is readily adaptable to automatic steering. Either gyrocompasses or magnetic compasses can be equipped for automatic steering, which keeps the vessel on course without hand operation.

Anchor machinery. The anchor gear, consisting of anchor windlass, chain stoppers, and hawsepipes (for stockless anchors), must be capable of paying out and stopping the chain, securing it, and later retrieving and stowing the anchor in the hawsepipe. The windlass rotates two wildcats (sprockets) that engage the anchor chains for port and starboard anchors.

The steam engine with reduction gearing was initially used for power-driven windlasses, and is still used on most vessels having a convenient source of deck steam. Direct electric drive with reduction gearing may be used on medium-sized vessels. Electro-hydraulic drive predominates, however, especially on large vessels. This consists of an electric motor driving a variable, reversible-stroke pump piped to a similar hydraulic unit, usually with fixed stroke, which serves to drive the wildcat through reduction gearing.

Windlasses ordinarily raise one anchor from 60 fathoms depth of water, or two anchors from 30 fathoms depth, at 5–6 fathoms/min (1 fathom = 6 ft).

Due to hawsepipe friction and other losses, the steam or electric power for mechanical drive is roughly twice the hoist horsepower of the anchor and overboard chain weight. The power required for hydraulic drive is approximately three times the hoist horsepower.

Hand-operated brakes are provided for stopping the chain when it is let go, unrestrained, in water not deeper than about 15 fathoms. For greater depths, the anchor is "walked out" slowly under power before being "let go."

When the chain has been payed out and stopped, it is secured by a chain stopper which is located between the windlass and hawsepipe and jams down against the chain, preventing further paying out while the vessel is at anchor. [E.F.H.E.]

Bibliography: D. Arnott (ed.), *The Design and Construction of Steel Merchant Ships*, 1955.

Refrigeration and heating. Standard refrigerants used on ships are Freon 12 and Freon 22. A type of compressor using Freon 12 in the refrigeration cycle is indicated in Fig. 13. After compression, the hot gas passes into a shell-and-tube type condenser where the circulating water absorbs heat, condensing the gas to a liquid which passes on to a receiver. The high-pressure liquid refrigerant then passes from the receiver through the liquid line, when the expansion valve allows flow, to the low-pressure expansion coils, where it evaporates. In vaporizing, the refrigerant absorbs the heat of vaporization from the evaporator coils, thus producing the refrigerating effect in the coils and their surroundings. The compressor suction line is connected to the evaporator coils and the vaporized refrigerant is drawn to the compressor to begin the cycle again.

Since different products require different temperature ranges for proper refrigeration, there are commonly at least three refrigerated compartments for ship's stores—a meat room at 0–10°F, a fruit and vegetable room at 40–50°F, and a dairy room at 35–50°F. The evaporator coils are usually located in the refrigerated compartment, the refrigerating effect being by direct expansion. In some cases an indirect system is used, cooled brine or water being circulated through pipe lines in the compartment. Ships carrying refrigerated cargo

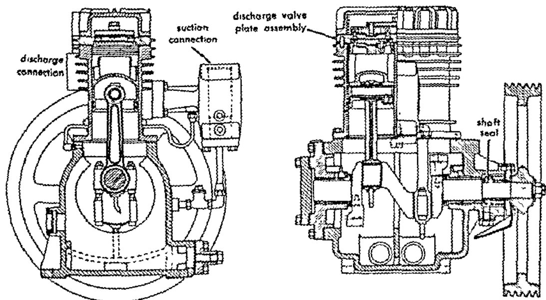


Fig. 13 Freon 12 compressor. (Society of Naval Architects and Marine Engineers)

have much more exacting refrigeration procedures and much larger refrigeration capacity.

Piping. The operation of all main propulsion, auxiliary, and deck machinery, and hence the ship itself, depends upon the various piping systems. For purposes of design and construction shipboard piping is divided into two classes, depending on the range of pressure and temperature and the fluid carried. Class I includes all piping for: (1) steam and air over 150 psi or over 370°F; (2) water over 150 psi or over 300°F; (3) oil over 150 psi or over 150°F; and (4) lethal gas and liquids at all pressures and temperatures. Class II includes all piping intended for working pressures and temperatures below those for Class I. For merchant vessels the regulations of the classification societies, as well as government regulatory safety rules, must be followed in selecting materials for piping and fittings, methods of manufacture, and methods of flange attachment. Thermal expansion must be allowed for by use of expansion bends or other suitable means.

ELECTRIC PLANTS

Shipboard electric plants have developed so far beyond their original purpose of simple electric lighting installations as to become extensive and complicated systems, essential to the operation and safety of ships and their personnel. The electric plant usually includes, in addition to two or more generators and the many motor-driven auxiliaries, the interior communication system, radio transmitters and receivers, radar and other navigational equipment, galley and hotel electrical equipment, motion-picture equipment, and the public-address system, as well as detecting, alarm and control equipment. The operation of the ship is so dependent on the continuing operation of the electric plant that emergency sources of electrical power, such as diesel generators and batteries, are normally provided. Repair parts must be kept aboard ship for each generator, motor, and electrical device. Generators must have close speed- and voltage-regulation if the varied equipment is to operate properly.

Many rules and regulations apply to American merchant ship electric plants, including those of the U.S. Coast Guard, the Federal Communications Commission, and the American Bureau of Shipping. In addition, vessel specifications usually require conformance with the American Institute of Electrical Engineers recommended practices on shipboard, as well as the regulations of the National Board of Fire Underwriters.

Electrical systems on ships may utilize generating plants of any of the following types, according to needs: (1) 120-volt, 2-wire dc; (2) 120/240-volt 3-wire dc; (3) 240-volt, 2-wire dc, with separate 120-volt, 2-wire dc motor-generators for lighting supply; (4) 230-volt, 3-phase, 3-wire ac; (5) 450-volt, 3-phase, 3-wire ac; and (6) a combination of types (4) or (5) with ac propulsion generators.

Direct-current plants are used where variable speed is important, since ac motors lack the torque and speed control characteristics which are available with dc motors. System 1 fits the needs of a vessel with small electrical load, 75 kw or less. System 2 supplies power at 115/230 volts and has the advantage that power distribution at 230 volts requires only about half the weight of copper required for 115 volts. Power for electric motors and galley and heating appliances is supplied at 230 volts, while power for lighting, fans, and portable appliances is at 115 volts. System 3 has the advantage of separating the lighting and power loads, though first cost is higher than for system 2. All ac generators require a voltage regulator and a dc exciter for field excitation. These items make the total cost of alternators about the same as dc generators of equal power, though the cost of maintenance of alternators is less. Systems 4 and 5 offer the advantages of power distribution at higher voltage. System 6 is used on some bulk carriers where the cargo handling load is much greater than the auxiliary electric load at sea, the latter load being handled by a system of type 4 or 5.

ARRANGEMENT AND MATERIALS

Arrangement of machinery. Before making a preliminary plan of the arrangement of machinery on a ship, it is necessary to make tentative selections of the various items and components of equipment which make up the plant as a whole and to know their approximate dimensions and particulars. With this information in hand, several acceptable arrangements may be possible, the problem of the designer being to select the best compromise. The preliminary plan usually becomes a part of the contract plans between the owners and the shipbuilder. Final plans are made from corrected drawings and particulars of the units actually selected for installation. Since space aboard ship is always at a premium it is important that the arrangement selected be as efficient and practical as possible. Guiding principles include (1) selection of equipment compatible with space limitations; (2) convenience of operation and easy accessibility for repairs and replacement of parts with minimum disturbance to piping or other units; (3) simplified piping and connections; (4) location of units which work together so that piping or other connections are short and direct; (5) location of equipment in accordance with functional characteristics; and (6) provision of safety guards in hazardous areas, such as casings around moving parts, insulation on hot surfaces, and handrails.

Materials used. The choice of materials for best design and construction demands care since many technological advances require either improved materials or more discerning use of those available. Thus much attention is given to improving materials or developing substitutes or combinations which are capable of performing the function more efficiently and, if possible, more cheaply.

Metals, both ferrous and nonferrous, are the most commonly used materials in marine machinery. Iron, copper, lead, and tin, in both impure and alloy form, had been in use on shore for centuries before their use in marine machinery. Zinc, nickel, aluminum, and magnesium are relatively new, though alloys containing zinc, for example brass, are very old (*see* ALLOY). Many nonmetallic materials are also used in marine machinery, including asbestos, cork, mineral wool, hair felt, fiberglass, diatomaceous earth, paper, rubber, mica, varnish, lacquer, shellac, paraffin, paint, brick, refractories, glass, plastic, wood, leather, solvents, abrasives, adhesives, and lubricants, with many variations in practically every item named.

Iron and steel, in their many forms and alloys, offer a wide range of such properties as strength in tension, compression and shear, fatigue strength, hardness, ductility, malleability, elongation, yield point, modulus of elasticity, notch sensitivity, and resistance to corrosion and erosion. Nonferrous metals and alloys as well as nonmetallic materials also offer a wide range of properties. Experience has established the general use of certain materials for certain purposes, though changes in practices are constantly occurring.

Corrosion is a major problem on all ships. Since it involves an electrochemical reaction between a metal and water, it is desirable to keep the water or moisture away from the metal insofar as feasible. Thus corrosion may be minimized by applying a coating of protective metal such as zinc (galvanizing), tin, lead, nickel or copper, or by the more common and generally used method of applying protective paints. Another factor in corrosion is the varying tendency of different metals to dissolve in water. Since the tendency varies, the metals may be arranged in a list (electrochemical series) of decreasing tendency (*see* ELECTROCHEMICAL SERIES). Electrolytic contact between dissimilar metals should be avoided, since it will stimulate corrosion. The extent of the dissimilarity, which is indicated by the relative position of the metals in the electrochemical series, is a factor in the rate of corrosion, though other factors may have a greater effect. Other methods of inhibiting corrosion include the use of a self-sacrificing metal like zinc to protect iron, elimination of oxygen in water (as in the case of boiler water), the use of corrosion-resistant alloys, and cathodic protection. The latter method involves the application of an external voltage which renders the structure to be protected cathodic and concentrates corrosion on auxiliary anodes used for the purpose. This method is being applied in the original design of some new ships and has been used effectively in protecting laid-up ships. *See* CORROSION; DRYDOCKING.

[K.K.C.]

Bibliography: J. M. Labberton and L. S. Marks (eds.), *Marine Engineers' Handbook*, 1945; H. L. Seward, *Marine Engineering*, vol. 1, 1942, vol. 2, 1944.

Marine marsh

A flat, vegetated land surface at the edge of the sea. This type of marsh is found on all coasts of the world and constitutes a discrete ecological unit inhabited by a characteristic flora and fauna. The physiography of the marsh depends on the geology and climate of the region, recent changes in sea level relative to the coast, the nature of the marsh vegetation, and the tidal regime. Tidal circulation serves to maintain aerobic conditions in the overlying water and mud surface and to transport biological nutrients and waste products into and out of the marsh. Geologically, a marsh is a transitory feature.

Development of a marine marsh proceeds in the following sequence. (1) Suspended material, mostly of terrestrial origin, is deposited along the edges of estuaries and in sheltered bays and lagoons. (2) When the surface of these deposits is built up above low-tide level, it is colonized by pioneer plants of the genus *Spartina* (cord grass or marsh grass) in the Northern Hemisphere (Fig. 1), and frequently by *Salicornia* (pickle weed) in the Southern Hemisphere. (3) Sedimentation is accelerated in the vegetated area, and the marsh continues to grow upward to the height of the highest tides. The rate of growth depends on the supply of suspended material and the total relief attained is equal to the maximum tidal range, including storm tides. (4) As the level of the marsh rises, elaborate drainage systems may develop. At the same time the frequency of the tidal flooding decreases, thus changing the physicochemical conditions and consequently the nature of the vegetation. On the higher elevations of Northern Hemisphere marshes *Spartina* is usually replaced by *Salicornia*, *Distichlis*, *Juncus*, *Panicum*, and other genera, depending on the geographical location (Fig. 2). This development is typical in temperate regions. In tropical regions, mangrove swamps are the counterparts of this type of marsh. Mangroves, however, become established in flat protected areas when the mud surface is still below low-tide level.



Fig. 1. View o[illegible] [illegible]sh on the coast of Georgia.

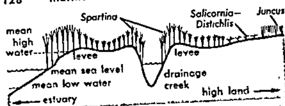


Fig. 2. Diagrammatic cross section of a marine marsh on the south Atlantic coast of the United States; vertical exaggeration about 10 to 1

In some marshes there are small, poorly drained areas called salt pans which develop as the sea water evaporates and the salt content of the mud increases. Pan areas are not invaded by the usual marsh plants

The number of different species found in marine marshes is relatively small when compared with other environments. Besides the larger rooted plants, the mud surface is inhabited by diatoms, dinoflagellates, green algae, and blue-green algae. A dense and varied bacterial flora is active in decomposing plant and animal remains. The more common invertebrate animals inhabiting marine marshes are various species of snails, mussels, and crabs. *See* ECOLOGY; ESTUARINE OCEANOGRAPHY; MANGROVE SWAMP; SWAMP, MARSH, AND BOG

[R.A R.]

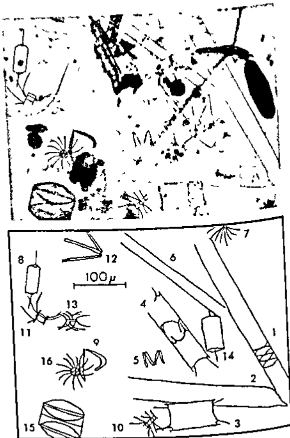
Marine microbiology

The study of the microscopic organisms living in the sea. Marine microorganisms were probably among the first (cellular) living entities and have played a tremendous part in geology and in the biology of the oceans. They are still of paramount importance as transformers of organic and inorganic substances and as food for other, often larger, organisms.

Marine microbes may be considered as (1) autotrophs and (2) heterotrophs or phagotrophs. Autotrophs live on inorganic materials and use carbon dioxide as the sole carbon source, obtaining their energy from sunlight (phototrophs) or from chemical reactions (chemoautotrophs). The heterotrophs and phagotrophs (bacteria, heterotrophic algae and fungi, and protozoa) require a source of carbon more complex than carbon dioxide; this they obtain, along with energy, by breaking down or ingesting other microbes, plants, and animals.

Phototrophic microorganisms. These constitute the most important plants in the sea, since all but the macroscopic seaweeds are microscopic and the latter constitute only about 2% of the plant life in the sea. The microscopic components are known as phytoplankton and occur in the upper waters of the oceans, the photic zone, where light can penetrate (*see* MARINE ECOSYSTEM). This zone varies in depth depending on the incident angle of sunlight and on the turbidity of the water. Most phytoplanktonic organisms are motile (for example, the flagellate protozoa), or asymmetrically shaped, or provided with processes (for example, the diatoms, see illustration) so that they tend to rotate in a way that brings them back towards the surface whenever they begin to sink. Many have oil as their storage product, which aids their buoyancy.

The phytoplankton is made up of a number of taxonomic groups, including diatoms and blue- [illegible] contain- [illegible] chla [illegible] hores, [illegible] chrysomonads, cryptomonads, and silicoflagellates (*see* COCCOLITHOPHORIDA). The organisms range in size from 1 millimeter long (a few diatoms) to 2 microns. In numbers, diatoms, blue-green algae, and the smaller flagellates are usually the most abundant. Diatoms are most numerous in boreal regions, blue-green algae are abundant in the tropical waters of the Pacific and Indian Oceans, coccolithophores and microflagellates in the warm and temperate waters, and dinoflagellates in inshore (neritic) waters. "Red tides" appear when some organisms locally grow to such a density that the water actually becomes colored; the reddish hue is due to accessory pigments. The Red Sea owes its name to the heavy blooms of a blue-green alga,



Marine phytoplankton, showing a number of species of diatoms. (1,6) *Rhizosolenia* sp.; (2) *Rhizosolenia alata* var. *indica*; (3,4) *Biddulphia chinensis*; (5) *Thalassiothrix nitzschioides*; (7,16) *Bacteriastrum varians*; (8) *Ditylum brightwellii*; (9) *Chaetoceros coarctatum* (fragment); (10,11) *Chaetoceros* sp.; (12) *Thalassiothrix frauenfeldii*; (13) *Chaetoceros laeve*; (14) *Ditylum sol*; (15) *Hyalodiscus stelliger*.

Oscillatoria (Trichodesmium) erythraea, which also occurs in huge masses in the Indian Ocean and tropical Pacific waters. Dinoflagellates form red tides in coastal waters and often contain a neurotoxin which may poison fish and other marine animals, causing heavy mortality. Such dinoflagellate blooms may produce large deposits of dead organic matter, as at Walvis Bay, South Africa; these have been considered to represent a possible source of petroleum, resulting from the bacterial reduction of the organic matter to hydrocarbons.

Chemoautotrophs. These are important only on the sea floor. They utilize energy-producing (exothermal) reactions such as the aerobic oxidation of sulfides (*Thiobacillus*), of ammonia (*Nitrobacter* and *Nitrosomonas*), or of hydrogen (*Hydrogenomonas*), and the anaerobic oxidation of hydrogen coupled with the reduction of sulfates to sulfides (*Desulfovibrio*). Thus they are important in the sulfur and nitrogen cycles of the sea bed, but are less common in the sea itself, except in shallow waters. In estuaries, bacteria of the sulfur cycle influence and sometimes control the biology of the overlying water.

Heterotrophs. While some diatoms and many dinoflagellates can live either as autotrophs or heterotrophs, and many flagellates are obligate heterotrophs, lacking chlorophyll, the most important heterotrophs in the sea are bacteria. These aid in the decomposition of organic matter in the sea and return much of this organic matter to the inorganic state, releasing carbon dioxide, nitrogen, phosphate, sulfur, and other essential elements in a form available for use by photosynthetic plants. Even refractory substances, such as cellulose and chitin, are decomposed by marine bacteria. Heterotrophic bacteria are not numerous in the open sea, except in the vicinity of plankton swarms, and their scarcity may be due to grazing by animals.

Phagotrophs. Phagotrophic protozoa ingest other phytoplankton organisms, bacteria, or detritus, and behave as do the other animals of the zooplankton. They include some flagellates, the tintinnids and other ciliates, and the foraminifera and radiolaria (rhizopods).

Interrelationships of microorganisms. Life in the sea is a vast symbiosis or metabiosis, and no organisms can exist independently of others. Moreover, the sea is never without microorganisms even in a particular region. There are always enough nearby to start another bloom if conditions are favorable. Many phytoplankton organisms require growth factors, and it is believed that these are frequently supplied by marine bacteria; some phytoplankton organisms produce toxic substances which inhibit the growth of other organisms. Bacteriophages have been detected in the Black Sea. Phosphate is accumulated and stored by diatoms and is released by bacterial action after their death; some types of phytoplankton (for example, the diatom *Fragilaria*) require nitrates, others (for example the blue-green algae) prefer or require ammonia, and a few can fix atmospheric nitrogen

Role of microorganisms. Marine microorganisms are the food of a large number of marine animals including many crustacea of the zooplankton, mollusca, and plant-eating fish. The phytoplankton that can be counted or measured represents merely the residue that is left after the animals have grazed. Thus, phytoplankton forms the primary food in the sea. Zooplankton animals appear to be selective in their feeding and to reject certain diatoms such as *Chaetoceros* and *Rhizosolenia*. *Chlorella* may be ingested by cyprid larvae but passes whole into the feces, and the larvae die on this diet. Selectivity in grazing and facultative heterotrophy of some phytoplankton organisms create difficulties in assessing phytoplankton production.

Many plankton microbes have calcareous or siliceous skeletons. Foraminifera and coccolithophores have, as a group, calcareous skeletons, and diatoms, radiolaria, and silicoflagellates have siliceous skeletons; these may accumulate on the sea bottom to form vast beds which may persist through geological ages, for example, foraminiferal limestones, diatomaceous oozes, and radiolarian jaspers.

Bacteria, and possibly diatoms, live on the sea floor, even at the greatest depths (10,000 meters), with hydrostatic pressures up to 1000 atmospheres. This bottom (benthic) flora forms a biosphere very distinct from the phytoplankton biosphere. See Algae; Bacteriophage; Ecologic interactions; Nitrobacteraceae; Protozoa; Sea water; Sea water fertility; Spirillaceae.

[E.J.F.W.]

Bibliography: E. J. F. Wood, Significance of marine microbiology, *Bacteriol. Rev.*, 22:1–19, 1958.

Marine navigation

The process of directing the movements of water craft from one point to another. For the three principal methods of navigation see Celestial navigation; Dead reckoning; Piloting.

Piloting is concerned primarily with locating positions of the vessel by means of landmarks or prominent features of the underwater topography. Until recent years it could be used only when the vessel was near land. With the development of various means of positioning a vessel by electronics, piloting has become available over much of the ocean areas of the world.

Dead reckoning is the determination of the position of a vessel by projecting a new position from a previously determined position on the basis of subsequent motion of the vessel. Its accuracy depends upon the accuracy of determining the direction and distance traveled.

Celestial navigation is the determination of the position of a vessel with the aid of celestial bodies. Observation of altitude of the body is made by a marine sextant. Comparison of the corrected sextant altitude with computed altitude for an assumed position provides information for determining a line of position. Two or more lines of position establish a fix.

Most charts used by mariners are on the Mercator projection, but other projections are also used. [A.B.M.]

See NAVIGATION.

[illegible] *American Practical* [illegible]

1958.

Marine refrigeration

The application of refrigeration on shipboard for the preservation of perishables in transit, or for the storage of food in ships' stores for the crew and passengers of merchant vessels and military craft Completely insulated reefer vessels are used for the transportation of chilled or frozen cargo. All oceangoing vessels have insulated ships' stores, and many merchant vessels have a few insulated cargo holds. Small refrigeration systems usually employ refrigerant 12 (Freon 12) and large installations refrigerant 11 (Freon 11) with brine circulation. Refrigerants 717 (ammonia) and 744 (carbon dioxide) are obsolete for American vessels. U.S. Coast Guard, U.S. Public Health Service, and American Bureau of Shipping regulations apply. *See* MARINE MACHINERY; REFRIGERATION [H.M.HE.]

Marine resources

The oceans cover 71% of the earth's surface, to an average depth of 3795 meters, with a total volume of 1.37×10^9 km^3. Their living and nonliving contents constitute the basis of several extractive industries. Many of the sea's resources, however, cannot be used profitably at the present stage of knowledge, but a moderate and reasonably certain advance in technology would make them valuable. The development of these latent resources is an important frontier of modern science. Both extractive and nonextractive resources of the oceans are discussed.

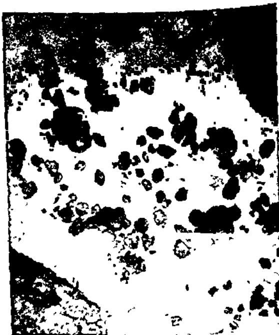
Extractive resources. These include (1) nonrenewable resources, or resources for which the rate of renewal is so slow as to be negligible, such as petroleum and natural gas under the sea floor, mineral deposits on the ocean bottom, dissolved minerals in the water, and the water itself; and (2) renewable resources, such as the living resources of the sea.

Petroleum and natural gas. The continental shelves (the land submerged under less than 600 ft of water) under the margins of the seas extend over about 11,800,000 square miles and include some 30,000,000 cubic miles of possible oil-bearing sediments. By comparison with the petroleum content of such sediments on land, it is estimated that they contain about 400,000,000,000 barrels of recoverable crude oil, plus large amounts of natural gas.

During the past decade, extensive geophysical and geological prospecting has located some of these deposits in the Gulf of Mexico, off the coast of California, in the Persian Gulf, and elsewhere. Successful drilling to recover them has been accomplished in depths of water up to 100 ft and at distances up to 40 miles from shore. Rapidly developing new techniques are extending the water depths and distances from shore in which drilling can be economically conducted. *See* MINERAL FUEL AREAS.

Minerals on the sea floor and in the water. The floor of the deep sea is known to contain low-grade deposits of cobalt, nickel, and copper (0.1–0.7% by weight of the metals) associated with deposits of iron and manganese. None of these are now utilized, but fairly large deposits of manganese nodules, discovered in 1957 on the tops of some Pacific seamounts, offer commercial possibilities.

Sea water itself contains a large variety of elements as dissolved salts. In 1,000,000 lb of sea water there are, for example, 18,980 lb of chlorine, 10,561 lb of sodium, 1,272 lb of magnesium, 380 lb of potassium, and 65 lb of bromine. Extraction of sea salt by evaporation is an ancient industry and is now highly developed both for the recovery of sodium chloride and for the production of sodium sulfate, potassium chloride, magnesium chloride, and *magnesium oxychloride cements.* Commercial *extraction of bromine from sea water* was initiated by the Ethyl Corporation in 1924, for the manufacture of the gasoline additive ethylene dibromide. The Dow Chemical Company initiated, in 1941 at Freeport, Texas, the production of magnesium metal from sea water, employing a combination of chemical and electrolytical processes.



Typical calcareous ooze with large nodules of manganese in the southeastern Pacific, latitude 42°50'S, longitude 125°32'W. Depth 4560 meters. (*Photo by C. Shipek, U.S. Navy Electronics Laboratory, San Diego, California*)

With sufficient cheap power, and with depletion of other sources of minerals, the production of some minerals from sea water will become increasingly feasible.

Living resources. The living resources of the sea support the largest, by far, of the extractive marine industries. The world's sea fisheries, producing protein products for human consumption and for animal foods and other purposes, yielded a catch in 1956 of 26,000,000 metric tons of fishes and marine invertebrates. Of this, 2,900,000 metric tons were landed in the United States. In addition, the annual catch of whales amounts to about 4,000,000 tons, mostly from the Antarctic.

Three-quarters of the fish catch is made in temperate waters of the Northern Hemisphere, mostly within a few hundred miles of the land, despite the fact that the southern oceans constitute 57% of the world's sea area. The disproportionately large yield from the Northern Hemisphere is related to three factors: (1) human populations are heavily concentrated there; (2) the major fishing nations are the industrialized maritime nations, which are located there; and (3) except for some tuna, salmon, and herring fisheries, the major sea fisheries are located in the relatively shallow areas along the continents, and the extent of these shallow areas is much greater in the Northern than in the Southern Hemisphere.

That the sea fisheries can yield greatly increased harvests is quite certain, because there are vast areas of the sea, especially in the Southern Hemisphere, which are scarcely fished at all, and because even in presently utilized areas there are a large number of known fish stocks which are not being harvested to their productive capacity. The sea-fish harvest has increased rapidly, from 17,000,000 metric tons in 1948 to 26,000,000 in 1956, but the protein deficit in many parts of the world will allow an even more rapid increase as technological advances in fish catching, processing, and distribution make possible the economic exploitation of the unused resources.

The plant life of the open sea consists of microscopic plants, phytoplankton, not amenable to direct harvesting. The larger seaweeds (algae) which are of commercial importance occur only along the shallow edges of the sea. The *Phaeophyta*, which include the giant kelps, are the basis of important industries in the United Kingdom, Japan, and the United States, since they contain a colloidal chemical substance, algin, similar to cellulose, which is of wide application in food and pharmaceutical products and in rubber and textile manufacturing. From certain genera of the *Rhodophyta*, or red algae, are produced agar and carrageenin. The world seaweed harvest amounted in 1956 to 370,000 metric tons and is capable of a large increase.

Despite the fact that the sea fishes are not being utilized anywhere near the limit of their biological potential, some scientists have become fascinated by the possibility of gathering organic material from the sea at a level lower in the food chain, such as zooplankton, which feed on microscopic plants and are, in turn, fed upon by the fishes and other higher organisms. It is, however, overlooked that, although the zooplankton production in the sea is large relative to that of the fishes, the standing crops of either represent a very small volume in a very large volume of water. The schooling habit of most commercial fishes makes it possible to catch them in economically feasible quantities. For the more dispersed zooplankton, which constitutes only a few parts in a million parts of water, the problem of profitably straining out the organisms from the water is formidable. At some future time, when man's protein food needs are more pressing than now and when new techniques may be developed, plankton harvesting may become commercially possible, but in the foreseeable future the food harvest must continue to depend mainly on the fishes. See MARINE FISHERIES.

Water. Fresh water, which is a critical resource in arid regions, may be recovered from the sea by distillation, ion-exchange, and other processes. Research and development studies of these techniques are receiving much attention, but a sizable cost reduction is yet required to make such water economically usable for agriculture on a large scale, although it is already feasible for some domestic and industrial uses.

Nonextractive resources. Other aspects of marine resources include disposal of waste products and use of the sea as a source of energy, as a medium of transportation, and for recreational purposes.

Waste disposal. Disposal of domestic sewage and industrial wastes is conveniently accomplished near coastal population centers by running them into the adjacent sea, where the large volume and rapid mixing of the waters dilute the wastes, and the bacteria in the sea break down the organic constituents. It is necessary, however, to consider in some detail the local effects of tidal and wind currents, density stratification, rates and volumes of mixing, the character of the bottom sediments, and the rate of disappearance of human bacteria as a basis for planning such disposal without incurring a pollution hazard

In the coming era of large nuclear fission power plants, it may be necessary, particularly for countries with densely populated land areas and long sea coasts, to dispose of some fission waste products in the sea. Safe ocean disposal of radioactive wastes involves the selection of sites where rapid and profound dilution will occur, or where sufficient decay will take place before the radioactive waters and their contained organisms come into contact with human beings. Some low-level radioactive wastes are already being safely disposed of in coastal waters, but it is certainly not safe to introduce large quantities of high-level wastes there. Deep-ocean disposal may be possible, but much more must be known about the deep-ocean circulation, and the transfer of elements between the

deep sea and the surface waters (where man uses the sea) by physical and biological processes, before it can be stated with certainty where and under what circumstances specified quantities of radioactive waste products can safely be introduced.

Energy. Of the several forms of energy in the sea that are capable of being used to produce power, the most apparent is the ebb and flow of the tide, and several attempts are being made to harness it, such as the Passamaquoddy project in the United States and the Race River project in France. A new French station has also been designed to derive power from the thermal gradient in the sea.

Transportation. Long-distance transportation of large cargoes by sea is the indispensable basis of international commerce. Major opportunities for more efficient use of this resource lie at the boundary between sea and land, since major problems and costs of ocean transportation are involved with getting cargoes on and off the ships. Increased knowledge of the effects of waves, currents, and tides on dredged channels and structures can provide the basis of improved harbor design and development. The possibility of creating large artificial harbors by using nuclear explosives may make possible the creation of good harbors on sea coasts where none exist.

Recreation The recreational aspects of the sea are of importance to coastal populations in the temperate and subtropical regions, not only in providing healthful sports and satisfaction of man's curiosity and desire for beauty, but as the basis of large tourist and service industries. Important technical problems arise in rectifying conflicts between this use of coastal waters and other uses, such as commercial fishing, waste disposal, and oil drilling. [M.B.S.]

Bibliography: R. Revelle et al., *The Effects of Atomic Radiation on Oceanography and Fisheries*, Natl. Acad. Sci.—Natl. Research Council Publ. 551, 1957; M. B. Schaefer and R. R. Revelle, *Natural Resources*, 1959; H. U Sverdrup, M. W. Johnson, and R. H. Fleming, *The Oceans*, 1942; D. K. Tressler and J. M. Lemon, *Marine Products of Commerce*, 2d ed., 1951.

Marine sediments

The accumulation of mineral materials and organic remains on the ocean floor. Such sediments show a wide range of characteristics and depth of deposition because of differences in their source, in the physicochemical processes operating in the marine environment, and in distances from land. The sampling and study of sediments on the sea floor is an important phase of oceanographic research and together with the study of sedimentary processes on land constitutes the subdivision of geology known as sedimentology. See OCEANOGRAPHY; SEDIMENTATION (GEOLOGY).

ENVIRONMENTS OF DEPOSITION

The sediments of the ocean floor can be subdivided into three broad categories as follows: sediments in coastal and nearshore areas, sediments in the marginal zone around continents and islands, and sediments in areas of the ocean far away from the continents.

Littoral, bay, and lagoon sediments. These sediments are predominantly of inorganic origin, as the nearshore and shallow water location along the sea coast is favorable to receive land-derived or terrigenous material. They are generally coarser and better sorted than sediments of the open sea and may be stratified as a result of intermittent fluctuations of the nearshore transporting agents. Sands are common, although sediment types vary locally in accordance with the principal physical and biological controls. The study of coastal deposition is of geological interest for providing evidence of former shoreline configuration and the interpretation of ancient sedimentary rocks.

The coastal environment is characterized by active and intense transporting agents (currents, waves, and tidal fluctuations) which are particularly strong in the shallow waters and are capable of churning up and transporting the coarser material. Rivers entering the sea set up a transitional environment between marine and fresh water and often cause rapid deposition of the land-derived river material.

The source of sedimentary material is provided by local subaqueous erosion (waves and currents) of the coastline and shallow bottom, and more distant subaerial erosion (rivers, winds, and ice) of the land with fluvial and eolian transport of mate- [illegible]

form of the coast line, type of rock on the shoreline, topographic irregularities of the bottom, nearshore circulation of water masses, and tidal currents. The interrelation of these factors determines the ultimate sediment type for a given locality. Similarly, these factors control the type of organic remains (oysters, clams, conchs, and other forms) which are associated with the sediments. See COASTAL LANDFORMS; SHORE PROCESSES.

Littoral deposits. The littoral zone is that area immediately adjacent to the shoreline; in a classic sense, pertaining to the marginal zone of the sea, and most commonly, pertaining to the intertidal zone. In this zone intense wave activity and strong tidal currents are capable of eroding even the most firm rocks of the adjacent land. The sediments here are generally coarse-grained and of fairly uniform size. For example, coarse gravels and cobbles are associated with steep rocky headlands exposed to strong wave activity. Where rocky headlands are absent and the shoreline is represented by a flat coastal plain, the littoral zone is typified by the uniformly graded sands characteristic of most beaches. The sands of the beach are often transported great distances along the coast by waves which move at angles to the coast.

Bay deposits. Bays are generally defined as a recess in the shoreline or an inlet of the sea between

two capes or headlands. These environments are influenced as much by terrestrial as marine factors. Generally, the shore is sandy and the bottom is covered with a blanket of sediment that becomes increasingly finer with increased depth and distance from the sea inlet. The bay mouth is likely to be composed of a coarser bottom, varying from gravel to cobbles or rock, where waves and tidal currents are intense. Such a progression in particle size occurs under ideal conditions in a well-protected bay with minimum wave and current action. The shallow flats bordering bays often exhibit highly variable physical conditions (including temperature and salinity) which result in local differences in the sediment types. These vary from sands to tidal-flat muds. Where burrowing organisms, such as clams, live in abundance they frequently churn the sediments to such a degree that original sediment stratification is destroyed. Where rivers empty into relatively enclosed bays, causing lower salinities, extensive oyster beds can develop. Rivers also contribute large quantities of silt and clay which often fill the bay head and form numerous shoal areas. *See* ESTUARINE OCEANOGRAPHY.

Lagoon deposits. Lagoons are shallow and elongate water bodies, usually oriented parallel to the coastline, and separated from the ocean by offshore bars or islands of marine origin. The sediments commonly consist of stratified sands and muds resulting from frequent alternations between periods when storm waves wash over the islands and deposit sands and periods of relative calm when fine-grained muds are deposited. Lagoon sediments often resemble marine sediments in areas where rivers are absent, but where rivers are abundant they are similar to bay deposits. [G.A.R.]

Neritic sediments. The neritic zone includes the shallow parts of the ocean (continental shelf or platform) which border the continents and most islands. It begins at the line of low tide and slopes gently at less than 1° to its seaward limit where a distinct break in slope marks the beginning of the steep descent to the ocean bottom. The depth of the break in slope varies, averaging about 400 ft, and not exceeding 600 ft. In areas of active tectonic movement the topography of the shelf is highly irregular. Off the coast of California the narrow shelf and the continental slope are separated by a wide zone of very deep basins alternating with shallow banks, sometimes crowned by islands. Steep-walled submarine canyons cut the outer edge of the neritic zone in many places. The canyon heads sometimes reach back into the immediate vicinity of the coast. *See* CONTINENTAL SHELF AND SLOPE; SUBMARINE CANYON.

The neritic environment is to a large extent a zone of active bottom and tidal currents. In the shallow part the waves are capable of churning up even coarse-grained bottom deposits; in the deeper parts they may prevent the deposition of fine-grained suspended material. Wind and waves set up longshore currents capable of transporting coarse material over long distances along the coast. Water temperatures vary greatly both seasonally and areally. Fresh-water runoff from the land causes marked salinity changes in the nearshore zone. Complex circulation systems connected with the outflow from rivers and estuaries occur.

Sediments deposited in the neritic zone belong to one or more of five groups: (1) biogenic (derived from organisms and consisting mostly of calcareous remains); (2) authigenic (precipitated from sea water, for example glauconite, phosphorite, salts, anhydrite); (3) residual (weathered from underlying rocks); (4) relict (remnants of earlier environments of deposition); and (5) detrital (products of the weathering and erosion of continents, supplied by streams or coast erosion, such as gravels, sand, silt, clay).

In areas with a high rate of supply of material derived from the land the sediments are in general noncalcareous and consist of gravels, sands, silts, and clays. Calcareous material may be accumulated locally in the form of shell beds, oyster reefs, and the like, but it is generally of minor importance. Where the supply of land-derived material is low, deposits of the neritic zone are primarily calcareous, consisting of remains of organisms and precipitated lime muds. The calcareous group is partly biogenic and formed locally, partly detrital and of either local or distant origin.

The study of modern neritic sediments is of great interest for geology because the vast majority of ancient sedimentary rocks were deposited in this zone. The bulk of the world's petroleum is also found in ancient neritic sediments. It is debatable, however, whether the modern neritic environment is entirely representative of the neritic environments of the past.

Noncalcareous neritic sediments. These deposits consist primarily of land-derived (terrigenous) noncalcareous material. As the depth towards the outer edge of the shelf increases, the ability of the transporting agents to move any but the finest suspended material gradually decreases. Theoretically, this should result in a decrease in particle size towards the deep ocean. In some areas such a decrease has been observed and a succession of zones from sand to silt to clay occurs from the coast outward. Many factors, however, can complicate this basic situation. The outermost portions of most continental shelves are covered with coarse material that appears to be of pre-Recent age and to have been deposited in shallow water during a time of low sea level. Since clays of Recent age occur in the deep ocean basins, it has been assumed that this absence of fine material on the outer shelf is not due to the inability of suspended particles to reach this area, but to mechanisms that would prevent its deposition there, as for example long-period surface waves, internal waves, and tsunamis (waves associated with marine earthquakes). The question has not been answered adequately at the moment, and it is probable that in many places the absence of deposition on the outer shelf is indeed due to a shortage of sediment supply.

The morphology of the coastal zone can control the sediment distribution on the continental shelf. Between Cape Cod and Cape Hatteras on the Atlantic Coast of the United States the major streams are deeply entrenched (occupy drowned valleys) and virtually all sediment is trapped in the estuaries and lagoons behind the barrier islands lining the coast. Sedimentation in the neritic zone is negligible and sandy deposits formed during the post-Pleistocene rise in sea level are exposed on the shelf (Fig. 1*a*).

The opposite is found where rivers bring in an [illegible]

ing on the current pattern the distribution of the clay zone may be asymmetrical with respect to the river mouth (Gulf of Mexico off the Mississippi delta). In this case there is no coarse-grained sediment on the outer shelf. A more complex situation develops where fine-grained sediment by-passes parts of the shelf that are exposed to vigorous wave action. The suspended clay of the Orinoco river in northeastern South America is carried northward by marine currents. It bypasses the shelf east of the island of Trinidad, where the bottom is shallow and exposed to the long Atlantic swell, and comes to rest on the deep and protected shelf north of Trinidad, 500 and more miles from the point of origin (Fig. 1*b*).

Where sediment supply from the land is restricted at the present time, either calcareous deposits of biogenic origin or relict detrital sediments occur. In the last case the distribution of sediments is primarily controlled by events associated with the post-Pleistocene sea level rise. The continental shelf of the Gulf of Mexico west of the Mississippi delta exhibits a series of subparallel sand ridges

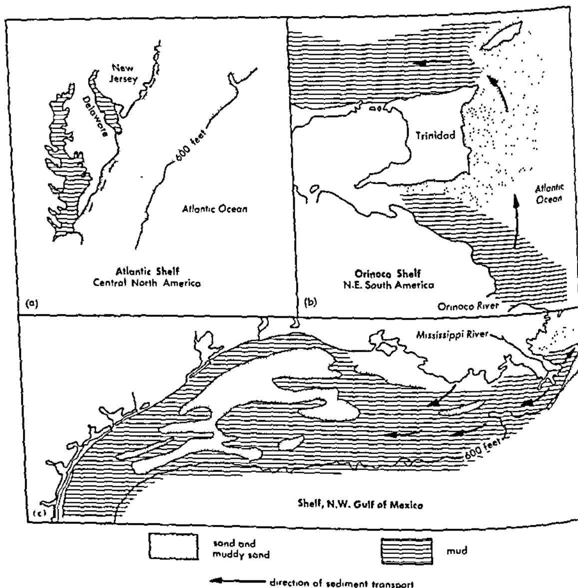


Fig 1. (a–c) Examples of sediment distribution in the neritic zone.

formed as shallow nearshore deposits during successive stages of the sea level rise. Modern sedimentation is restricted to the nearshore zone and is predominantly sandy as a result of extensive winnowing by waves. A little clay settles farther out in the neritic zone in the low areas between the old sand ridges (Fig. 1*c*).

In coastal areas having complicated topography associated with active tectonism, complex sediment patterns of relict, residual, biogenic, authigenic, and detrital materials occur. The pattern depends on the past depositional history of the area, the variations in vigor of the transporting agents which, in turn, are related to the topography, and the rocks outcropping on the sea floor. Under these circumstances, extreme and rapid fluctuations in the composition of the sediments can occur.

Very coarse-grained sediments are rare in the neritic environment. They are restricted to a narrow zone at the foot of coastal or submerged cliffs and cannot be transported far. The deposits in arctic and subarctic regions are an exception because very poorly sorted material ranging from large blocks to fine clay can be transported over considerable distances by ice breaking from glaciers.

Part of the suspended sediment never comes to rest in the neritic environment, but is carried into the deeper basins where it is deposited as pelagic clay on the continental slope and ocean floor. Sediments from the neritic zone can also be carried down into deep basins by turbidity currents flowing through submarine canyons. Where these canyons head in shallow nearshore waters, they tap the coarse-grained stream of sediment moving along the shore and permit transfer of sand and even gravel into deep water.

The distributions of neritic sediments discussed here are mere examples. The infinite variety of the numerous factors that control sedimentation in the neritic environment (currents, wind, waves, bottom topography, coastal morphology, abundance and type of sediment supplied, previous depositional history) results in an equally rich variety of sedimentary patterns. *See* DELTA.

Calcareous neritic sediments. These deposits are formed in the shallow parts of the ocean which border the continents and most islands, and in areas where the supply of land-derived sediment is of minor importance. In such areas, sediment is produced by biological activity in the form of the largely calcareous skeletons of animals, shells, Foraminifera, corals, and algae. The production of this material is highest in warm and clear waters, and calcareous neritic sediments are most abundant in tropical and subtropical areas. Wave action may break up and round the calcareous particles and skeletons into a calcareous sand. Lime muds are precipitated by the action of algae and perhaps also inorganically. In turbulent waters oolites, round sand-size particles, are formed by concentric precipitation of carbonate around a nucleus which may be a quartz grain. *See* ALGAE FOSSILS; CORAL; CORALLINE ALGAE; OOLITE AND PISOLITE.

Particularly important among the calcareous neritic sediments are reef deposits of various kinds, but large areas are covered with calcareous sediments that are not of reef origin, for example along the west coast of Florida. Calcareous sediments can also be produced when the sediment supply from the adjacent land consists primarily of limestone fragments. *See* GULF OF MEXICO.

Compared to the noncalcareous neritic deposits, calcareous sediments are of restricted extent, and with the exception of reef deposits, their formation and character are not yet well understood. *See* BARRIER REEF; CORAL REEF; FRINGING REEF; ORGANIC REEF. [T.V.A.]

Pelagic sediments. Nearly two-thirds of the crust of the earth is covered by pelagic sediments. These slowly accumulating oozes and clays which blanket most of the deep-sea floor are composed of the residue of planktonic life, wind-blown dust, cosmic dust, current-transported silt and clay, and authigenic minerals (Table 1). Beginning with deep-sea exploration in the mid-nineteenth century, the early deductive picture of a naked abyss was subsequently replaced by a concept (picturesquely termed the great snowfall) suggesting a gentle,

Table 1. Areas covered by pelagic sediments*

Type of sediment	Atlantic Ocean Area, $\times 10^6$ km^2	Atlantic Ocean %	Pacific Ocean Area, $\times 10^6$ km^2	Pacific Ocean %	Indian Ocean Area, $\times 10^6$ km^2	Indian Ocean %	Total Area, $\times 10^6$ km^2	Total %
Calcareous oozes								
Globigerina	40.1		51.9		34 4			
Pteropod	1.5							
Total	41.6	67.5	51.9	36.2	34.4	54.3	127.9	47.7
Siliceous oozes								
Diatom	4.1		14.4		12.6			
Radiolarian			6.6		0.3			
Total	4.1	6.7	21.0	14.7	12.9	20.4	38.0	14.2
Red clay	15.9	25.3	70.3	49.1	16.0	25.3	102.2	38.1
Grand total	61.6	100.0	143.2	100.0	63.3	100.0	268 1	100.0

* H. U. Sverdrup et al., *The Oceans*, Prentice-Hall, 1942.

downward drift of material from above slowly to form a veil of ooze and clay over the deep-sea floor. This concept has been modified by the discovery of strong deep-sea currents and turbidity currents which repeatedly reshuffle the pelagic sediments. These dynamic processes are important factors when considering the origin and source of pelagic sediments, rates of deposition, and the climatic record in deep-sea sediments. There are, however, areas where the "great snowfall" has not been significantly altered and where a fairly complete record of past events may be determined through a study of the variation of fossils and chemical properties.

Sediments of the ocean floor were early divided into terrigenous and pelagic. Terrigenous sediments are found in the vicinity of the continents and consist largely of products of land erosion which are transported to the sea by rivers. As distance from shore increases, the proportion of terrigenous material gradually decreases until the remains of pelagic life become the dominant components. Pelagic sediments generally are divided into abyssal clays and organic oozes. The two major types are abyssal red clay and globigerina ooze; the latter consists largely of fossil Foraminifera, mainly of the genera *Globigerina*, *Globigerinoides*, and *Globorotalia*. The percentage of 30% calcium carbonate is taken as the arbitrary dividing line between these two transitional types. In tropical and temperate latitudes, these two types are overwhelmingly dominant. In polar areas, diatoms dominate over the Foraminifera. Although the Foraminifera are the most obvious constituent of globigerina ooze, coccoliths (the remains of the coccolithophorids) make up a large proportion of the sediment. In some restricted areas, pteropod tests make up a large part of the carbonate fraction of the sediment. The physical composition of pelagic sediments and texture of mineral particles are shown in Table 2.

The distribution of pelagic sediments is closely related to depth and bottom temperature (Fig. 2). The red clays are found in the deepest, coldest areas, while the calcareous oozes are confined to the higher areas and to the basins filled with warm waters. The solubility of calcium carbonate is controlled by pressure, temperature, and CO_2 content. Red clay is considered a residue which remains when the carbonate of organic oozes is redissolved. The chemical composition of red clay is given in Table 3. [B.C.H.]

Bibliography: *Finding Ancient Shorelines*, Soc. Econ. Paleontol. Mineral. Spec. Publ. 3, 1955; J. W. Hedgpeth (ed.), *Treatise on Marine Ecology and Paleoecology*, Geol. Soc. Am. Mem. 67, vol. 1, 1957; P. H. Kuenen, *Marine Geology*, 1950; R. J. LeBlanc and J. R. Breeding (eds.), *Regional Aspects of Carbonate Deposition*, Soc. Econ. Paleontol. Mineral. Spec. Publ. 5, 1957; F. P. Shepard, *Submarine Geology*, 1948; F. P. Shepard and G. A. Rusnak, Texas bay sediments, *Inst. of Marine Sciences*, vol 4, no. 2, 1957; H. U. Sverdrup et al., *The Oceans*, 1942; P. D. Trask (ed.), *Recent Marine Sediments*, Soc. Econ. Paleontol. Mineral. Spec. Publ. 4, reprint, 1955.

Table 2. Physical composition of pelagic sediments and texture of mineral particles[a]

Characteristics		Red clay, %	Radiolarian ooze, %	Diatom ooze, %	Globigerina ooze, %	Pteropod ooze, %
Physical composition						
$CaCO_3$	Maximum	29.0	20.0	36.3	97.2	98.5
	Minimum	0	Trace	2.0	30.0	41.8
	Average	10.4	4.0	23.0	64.7	73.9
Planktonic Foraminifera	Maximum	27.0			95.0	75.0
	Minimum	0			15.0	15.0
	Average	8.8	3.1	3.1	58.9	34.7
Benthic Foraminifera	Maximum	3.0			10.0	10.0
	Minimum	0			0	Trace
	Average	0.6	0.1	1.6	2.1	3.6
Other calcareous remains	Maximum	6.3			26.0	57.0
	Minimum	0			Trace	15.8
	Average	1.0	0.8	3.2	3.7	35.5
Siliceous remains	Maximum	5.0	80.0	60.0	15.0[b]	20.0
	Minimum	0	30.0	20.0	Trace	Trace
	Average	0.7	51.4	41.0	1.7	1.9
Texture of mineral particles						
>.05-mm diameter	Maximum	60.0[c]	5.0	25.0	50.0[c]	20.0
	Minimum	Trace	1.0	3.0	Trace	Trace
	Average	2.4	1.7	15.6	5.1	4.7
<.05-mm diameter	Maximum	100.0	67.0[d]	27.9[d]	69.3	41.8
	Minimum	31.0	17.0[d]	12.5[d]	1.2	Trace
	Average	86.5	39.9[d]	20.1[d]	28.5	19.6

[a] After R. Revelle, 1936; adapted from H. U. Sverdrup et al., *The Oceans*, Prentice-Hall, 1942.
[b] Only in two exceptional cases; usual maximum <5%.
[c] Only in one exceptional case.
[d] Includes finely divided remains of siliceous organisms.

Table 3. Chemical composition of red clay (free of salt, organic matter, and $CaCO_3$)[a]

Constituents	%
SiO_2	54 48
TiO_2	0 98
Al_2O_3	15 94
Cr_2O_3	0 012
Fe_2O_3	8 66
FeO	0 84
NiCoO	0 039
MnO	
MnO_2	1.21
MgO	3 31
CaO	1.96
SrO	0 056
BaO	0 20
K_2O	2 85
Na_2O	2 05
V_2O_5	0 035
As_2O_3	0 001
MoO_3	Trace
P_2O_5	0 30
CuO	0 24
PbO	0 008
ZnO	0 005
ZrO	
H_2O	7 04
Total %	100 00
Total Fe as Fe_2O_3	9 59
Total Mn as MnO_2	1 21

[a] Adapted from H U. Sverdrup *et al.*, *The Oceans*, Prentice-Hall, 1942. Based on analysis by Steiger of composite sample from 51 representative red clay specimens

PHYSICAL PROPERTIES

Physical properties of marine sediments such as density and the elastic constants depend on many factors. These include grain shapes, sizes, and compositions; the amount of interstitial fluid and its properties; the nature of grain-to-grain contacts; the degree of compaction and consolidation; and the age. Measurable physical quantities appear to be affected to a much greater extent by the fractional volume of fluid in the sediment (porosity) than by sediment type. This is to be expected since density and elastic constants do not differ much for the principal constituents of sediments (silica, calcium carbonate, clay minerals). Water depth affects physical properties only to a slight extent.

Some measurements of physical properties are made on samples recovered from the ocean bottom by coring devices or in sufficiently shallow water by divers. Such observations are limited to sediments lying within a few tens of feet of the water-sediment interface. Properties of deeper-lying sediments are known from seismic refraction measurements of the velocities of elastic waves, by inference from dispersion of surface waves, or from gravity data. A few typical measurements are given in Table 4. These are merely illustrative and are not in any sense average or most likely values.

Definitions and some useful interrelationships among measurable quantities are the following:

$$\rho = \rho_1\phi + \rho_2(1-\phi) \quad \text{(bulk density)} \qquad (1)$$

$$\alpha = \sqrt{(k + \frac{4}{3}\mu)/\rho} \quad \text{(compressional wave velocity)} \qquad (2)$$

$$\beta = \sqrt{\mu/\rho} \quad \text{(shear wave velocity)} \qquad (3)$$

$$(\alpha/\beta)^2 = 2(1-\sigma)/(1-2\sigma) \qquad (4)$$

$$k = 1/C \quad \text{(incompressibility} = 1/\text{compressibility)} \qquad (5)$$

where ρ_1 and ρ_2 are fluid and average particle densities respectively, ϕ is the porosity ($\phi = 1$ at 100% fluid, $\phi = 0$ at 0% fluid), μ is rigidity, and σ is Poisson's ratio.

Several general conclusions may be drawn from observations. Density and porosity are nearly linearly related and for most ocean sediments observations lie between the two lines.

$$\rho_{cgs} = 1.03 + (1.67 \pm 0.05)(1-\phi) \qquad (6)$$

Table 4. Typical measurements of selected properties of marine sediments

Property measured	Fine sand, 17-station average[a]	Clayey fine silt[a]	Gray clay or silt[b]		Cream calcisiltite[c]		Gray clay[c]	Artificially compacted globigerina ooze pressure[d], kg/cm²		
								512	768	1024
Medium grain diameter, mm	0.19	0.02			[0.01][e]	[0.01]				
ρ, g/cm³	1.93	1.60	1.72	1.58	1.60	1.71	1.46	2.14	2.22	2.26
ϕ, %	46.2	65.6	[56]	[65]	65	57	74	[32]	[28]	[26]
α, km/sec	[1.68]	[1.46]			1.59	1.68	1.49	2.68	2.89	3.06
β, km/sec								1.20	1.42	1.57
σ (Poisson's ratio)	0 44[f]	0 50[f]						0.37	0.34	0.32
k, 10^{-11} dyne/cm²	[0.472]	[0.342]						1.13	1.25	1.38
μ, 10^{-11} dyne/cm²	[0.06][g]	[0 00][g]						0.31	0 45	0.56
Thermal conductivity, 10^{-4} cal/(cm)(°C)(sec)			26.8	23.1						
Approximate water depth, fathoms	15	15	1000	1000	2300	1600	2550			

[a] After E. L. Hamilton et al., 1956. [b] After E. Bullard, 1954. [c] After G. H. Sutton et al., 1957. [d] After A. S. Laughton, 1957. [e] Brackets indicate conversion of units from those used in the original publication. [f] Lower limit. [g] Upper limit.

Compressional wave velocities for $\phi > 0.6$ agree well with the predictions (equation) of A. B. Wood. This is obtained by inserting into Eq. (2) $\mu = 0$ and $C = 1/k = C_1\phi + C_2(1-\phi)$ where C_1 and C_2 are compressibilities of fluid and particles respectively. At smaller porosities compressional velocity rises more steeply, the $\phi = 0$ limit being near 6 km/sec. Poisson's ratio may be expected to vary from a value near 0.25 at zero porosity to 0.05 at $\mu = 0$. Good agreement of α with the Wood equation for $\phi > 0.6$ indicates that σ reaches its upper limit near $\phi = 0.6$

Seismic refraction measurements indicate that compressional velocity increases with depth in the sedimentary column, the gradient being from 0.5 to 2.0/sec. Thus ρ, μ, k, β should also increase with depth and ϕ and σ decrease. [J.E.N.]

Bibliography: E. Bullard, The flow of heat through the floor of the Atlantic Ocean, *Proc. Royal Soc. London*, ser. A, 222(1150):408–429, 1954; E. L. Hamilton, G. Shumway, H. W. Menard, and C. J. Shipek, Acoustic and other physical properties of shallow-water sediments off San Diego, *J. Acoust. Soc. Am.*, 28(1):1–15, 1956; A. S. Laughton, Sound propagation in compacted ocean sediments, *Geophysics*, 22(2):233–260, 1957; J. E. Nafe and C. L. Drake, Variation with depth in shallow and deep water marine sediments of porosity, density and the velocities of compressional and shear waves, *Geophysics*, 22(3):523–552, 1957; C. B. Officer, Jr., A deep-sea seismic reflection profile, *Geophysics*, 20(2):270–282, 1955; G. H. Sutton, H. Berckhemer, and J. E. Nafe, Physical analysis of deep sea sediments, *Geophysics*, 22(4):779–821, 1957; A. B. Wood, *A Textbook of Sound*, 1941

TRANSPORT AND RATE OF SEDIMENTATION

Transport of sediments. Detrital sediments are brought to the sea by rivers and streams. Most of this material is deposited in or near deltas or is swept along the beach by longshore currents and by the action of surf. A small proportion of the finest material is carried out to sea by the surface currents. Off some rivers, such as the Congo and the Amazon, a surface flow of muddy, brackish water extends several hundred miles seaward of the river mouth. Nevertheless the greater proportion of the sediment brought by streams to modern coasts remains in the coastal or deltaic environments.

The fine sediment which escapes the shore environment onto the continental shelf generally does not find a permanent resting place there but is either returned to the coast or is washed farther seaward by ocean currents until it reaches the quieter waters of the continental slope.

Sediments accumulate on the continental slope for a comparatively short interval until slumping, often triggered by earthquakes, carries the sediments down the continental slope. If the sediments in a particular area are relatively silty or sandy the slumps will transform (by mixing with water or simply by a loss of cohesive strength) and flow as turbidity currents. Thus sand, silt and lutite-sized detrital sediment eventually finds its way to the base of the continental slope. *See* SUBMARINE TOPOGRAPHY; TURBIDITY CURRENT.

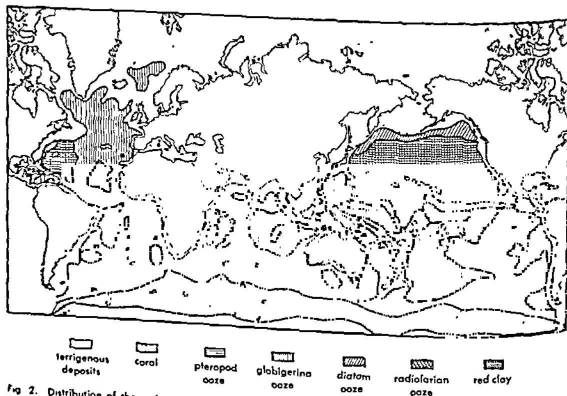


Fig. 2. Distribution of the major types of deep-sea sediments.

Turbidity currents, although of major importance in the modern seas, were much more important during the glacial epochs, when, as a result of the lowered sea level, river mouths generally lay near the outer edges of the continental shelves. Sediments carried to the river mouths in the beds of these streams were transformed directly into turbidity currents which flowed to the greater depths of the ocean basin carrying vast quantities of silt, sand, and gravel to the abyssal plains. These turbidity currents may have eroded submarine canyons, built the leveed midocean canyons, and, by spreading their deposits out on the ocean floor, may have formed the great submarine alluvial cones and vast abyssal plains.

Fine sediment is eroded by ocean currents from the tops and sides of the sharper and higher peaks of submarine mountains leaving lag deposits of coarse sand and gravel on the rocky bottoms. The finer material eroded from these mountains is deposited in their lee in veritable submarine "snow drifts" of sediment.

In the higher latitudes floating ice accumulates sediment along the shore, later floats out to sea, melts, and deposits its load. The transport of sediment by sea lions, penguins, and kelp is similar but of minor importance.

Over the ages a great mass of sediment is carried to the depths of the sea by the gradual sinking of (1) discarded shells of planktonic organisms, (2) fine detritus carried out from land in the nearly clear oceanic waters, (3) volcanic ash, dust, and even sand blown seaward by the wind, and (4) cosmic dust. In the equatorial Atlantic midoceanic area the combined contribution from these sources is about 1.7 g/(cm²)(1000 yr) at present rates; but during the last glacial stage the rate of pelagic deposition in the same area exceeded 3.5 g/(cm²)(1000 yr).

Rates of sedimentation. The earliest estimates of the rate of deep-sea sedimentation were based on the apparent depth of burial of deep-sea cables which had remained on the sea floor for a quarter of a century. Since cables may gradually sink into the sediment, and furthermore, since there is no sure way of determining the depth of burial, these estimates were worthless.

The second method was based on the correlation of the climatic record in sediment cores with dated events observed in other environments. By a study of temperature-sensitive pelagic Foraminifera, C. Schott was able to recognize the end of the last glaciation in short sediment cores obtained in the equatorial Atlantic. At that time (1935), Pleistocene geologists generally accepted 20,000 years as the duration of the Recent. Using this figure, he calculated rates of sedimentation. However, it is now known that the Recent was only about half as long as Schott assumed and that his rates of sedimentation were too small.

The ionium disequilibrium method of determining rates of sedimentation is based on the inequilibrium relationships of the radioelements of the uranium-235 series. Theoretically applicable in the time span from the present to about 300,000 years ago, this method depends upon the decay of ionium to equilibrium with its parent uranium (Fig. 3). The initial disequilibrium occurs because ionium is more rapidly removed from sea water than uranium. The method has numerous assumptions including the constancy with time of both the rate of deposition and the composition of the sediments. Reliable determinations of rates of sedimentation by this method are rare. Although many workers believe the method holds future promise, at present it is considered reliable only in abyssal red clay. Since this material is nearly unfossiliferous and cannot be dated by other methods, the rates determined are not subject to verification and do not allow the dating of past events. By far the most satisfactory method of age determination in deep-sea sediments is radiocarbon analysis. The radiocarbon locked in the biogenetic carbon or carbonate gradually decays in the deposited sediment. The method is, in general, only applicable in the range from the present to 40,000 years ago, but may be extended to twice that figure by current advancements in technique. Unlike the ionium method, it works best on fossiliferous sediments and therefore allows the dating of paleontologically determined events. Several other radioactive methods are under study but none of these have yet had much success. See RADIOCARBON DATING.

Early dating of glacial chronology was based on the counts of annual laminae (varves) in glacial lakes. Under certain special conditions, varves are formed in anaerobic marine basins. In these special cases, varve counting is a useful way to determine rates of sedimentation in marine sediments. See VARVE.

Rates of deposition. Measured rates of deposition of recent pelagic sediments (Fig. 4) range from 0.01 cm/1000 yr to over 60 cm/1000 yr. Rates of deposition of abyssal red clays determined by the ionium method range from 0.07 to 0.20 cm/1000 yr. The recent rate of deposition of globigerina ooze in the central equatorial Atlantic is approximately 1.6 cm/1000 yr. Rates of sedimentation are greater near the continents where terrigenous sedi-

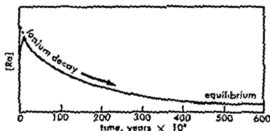


Fig. 3. The ionium method for measuring rates of deposition is based on this theoretical curve which indicates how the ionium decays. The method assumes that the rate and composition of sedimentation are constant with time. (After H. L. Volchok and J. L. Kulp, Geochim. et Cosmochim. Acta, 11:221, 1957)

ments constitute a greater proportion of the sediment. A rate of 63 cm/1000 yr has been determined in one core from the upper continental rise off Maryland.

Variation with topographic location. Rates of sedimentation vary rapidly from place to place because of the effects of currents, scour, slumping, creep, and turbidity currents. The fact that sub-

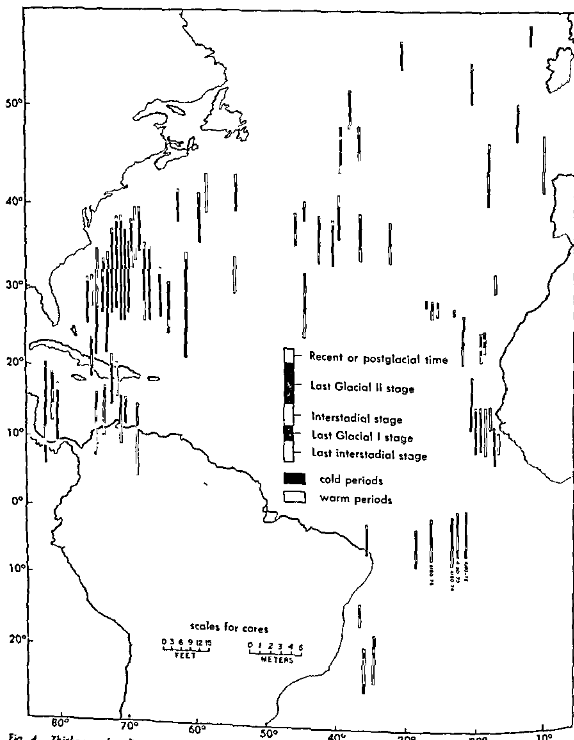


Fig. 4. Thickness of sediments laid down during past climatic stages. Top of column is location at which core was obtained. Stratigraphy is based on temperature-sensitive planktonic Foraminifera. Duration of the Recent has been determined as 11,000 years by radiocarbon measurements; however, rates determined for postglacial time cannot be extrapolated backwards in time since glacial rates of deposition were larger than postglacial rates. (After D. B. Ericson, M. Ewing, G. Wollin, and B. C. Heezen, Geol. Soc. Am. Bull., vol. 71, 1960)

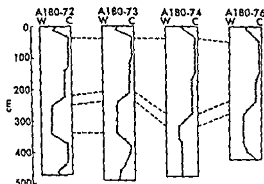


Fig. 5. Climatic curves for a profile of four equatorial cores. (After D. B. Ericson and G. Wollin, Correlation of six cores from the equatorial Atlantic and the Caribbean, *Deep-Sea Research*, 3(2):104–125, 1956)

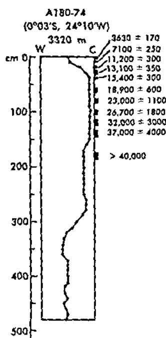


Fig. 6. Radiocarbon dates on core A180-74 of the series of cores shown in Fig. 5. (After W. S. Broecker, K. K. Turekian, and B. C. Heezen, *Am. J. Sci.*, 256(7): 503–517, 1958)

bottom echoes observed on echograms indicate that the upper layers of sediments thin over highs and thicken in depressions is probably to be explained as the effect of gentle bottom scour by ocean currents.

Variation with time. Early workers generally assumed that the rates of deposition of certain components of the sediment were, at a given location, constant with time. It was even suggested that by measuring the cumulative mass per unit area of a nonbiogenetic lutite (shale or clay) component in a sediment core, ages could be assigned, provided, of course, that the rate of deposition of the component was determined at some point by another method. A sufficient number of sediment cores have been radiocarbon-dated at a number of levels to demonstrate conclusively that the rate of deposition varies significantly with time even in purely pelagic deposits.

A study of the variation of rates of deposition with climate in one group of cores from the mid-equatorial Atlantic showed that both the carbonate and clay fractions underwent an abrupt decline in sedimentation rate at the end of the last glaciation. The glacial to postglacial ratio of sedimentation rate in the area is 3.7 for the clay fraction and 2.1 for the carbonate (Figs. 5–9). [B.C.H.]

Bibliography: W. S. Broecker, K. K. Turekian, and B. C. Heezen, The relation of deep-sea sedimentation rates to variations in climates, *American J. Sci*, 256(7):503-517, 1958; D. B. Ericson, M. Ewing, G. Wollin, and B. C. Heezen, Two hundred and twenty one Atlantic deep sea cores, *Geol. Soc. Am. Bull*, vol. 71, 1960; D. B Ericson, M. Ewing, B. C. Heezen, and G. Wollin, *Sediment Deposition in the Deep Atlantic*, Geol. Soc. Am. Spec. Paper 62, 1955; B. C. Heezen, M. Tharp, and M. Ewing, *The Floors of the Oceans: 1, The North Atlantic*, Geol. Soc. Am. Spec. Paper 65, 1959.

CLIMATIC RECORD IN SEDIMENTS

The climatic history of the ocean is recorded in layer upon layer of fossil planktonic animals and plants recovered in deep-sea sediment cores. The record is not everywhere preserved; layers are missing, repeated, and mixed. Turbidity currents, bottom scour, slumps, and postdepositional solution all operate to destroy the record, with the result that no more than half of all deep-sea sediment cores contain a reasonably complete and uncomplicated record.

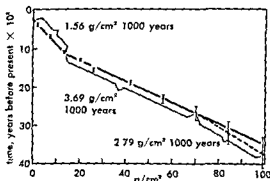


Fig. 7. Cumulative curve of total weight of salt-free dry material as a function of time in core A180-74. The slopes of the straight-line segments, which fit to the points, are directly related to sedimentation rate. Note the abrupt change in sedimentation rate which coincides with the end of the last glacial stage about 11,000 years ago. (After W. S. Broecker, K. K. Turekian, and B. C. Heezen, *Am. J. Sci.*, 256(7):503–517, 1958)

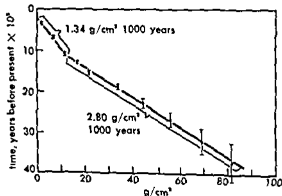


Fig. 8 Cumulative curve of weight of carbonate as a function of time before present in core A180-74. Rate of deposition of carbonate during glacial time was markedly higher than in postglacial time (*After W. S Broecker, K. K. Turekian, and B C. Heezen, Am. J. Sci., 256(7):503–517, 1958*)

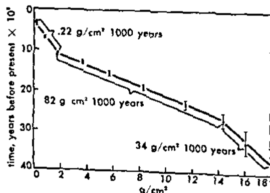


Fig. 9. Cumulative curve of weight of "clay" as a function of time before present in core A180-74. Rate of clay sedimentation during the glacial time was four times that in postglacial time. Due to perfect correlation of the line of equatorial cores shown in Fig. 5, results shown in Figs. 7–9 can be taken as typical of equatorial Atlantic eupelagic sediments. (*After W. S. Broecker, K. K. Turekian, and B. C. Heezen, Am. J. Sci., 256(7):503–517, 1958*)

Methods. The climatic history of the oceans is primarily based on the fossil Foraminifera. Several of the planktonic species have definite temperature tolerances. Thus, from a knowledge of their distribution in modern seas (Fig. 10), past climates may be inferred from a study of variations of abundances of these species in a core. Small samples of sediment, taken at frequent intervals in the core, are washed on a screen to remove the clay- and silt-sized particles (particles smaller than 74μ). The remainder, largely composed of Foraminifera, is examined under the microscope. The relative abundances of a number of species of Foraminifera are then tabulated (Table 5). From such a table a climatic curve is drawn (Fig. 11*a*). *See* FORAMINIFERA FOSSILS.

The discovery that the O^{18}/O^{16} ratio in biogenetic carbonate varies with the temperature of the water in which the shell grew provided an additional method for the study of paleoclimates. Although difficult corrections make the determination of absolute temperatures impossible, the results provide reliable indications of climatic fluctuations that are similar to those obtained by a study of foraminiferal tests. *See* GEOLOGIC THERMOMETRY.

It was formerly believed that small variations observed in the carbonate content of deep-sea cores could be directly related to temperature variations. It was assumed that temperature changes had produced variations in productivity of lime-secreting organisms. Radiocarbon determinations of rates of deposition, however, showed that in several cores, carbonate production had been greater at lower temperatures. It is now known that the rate of deposition of all major components in deep-sea sediments has varied widely in time. *See* PALEOECOLOGY (GEOCHEMICAL ASPECTS).

Nature of the climatic record. G. Schott identified the glacial to postglacial transition in many relatively short cores obtained by the *Meteor* in the equatorial Atlantic. Subsequently D. Ericson, Phleger, Todd, Menzies, and others worked out the late Pleistocene and Recent climatic record in many, much longer piston cores taken primarily from the Atlantic. C. Emiliani, working with oxygen isotope ratios, arrived at temperature fluctuations similar to those previously worked out by Ericson on a faunal basis (Fig. 11*b*). Although the final abrupt change from cold to warm is unequivocally identified by all workers as the end of the Wisconsin or Würm glacial stage, the earlier fluctuations have been correlated in a variety of ways. Part of the difficulty arises from the inaccuracies of the standard glacial chronology.

Many geologists formerly assumed that Pleistocene climatic changes had occurred gradually, with the warm and cold periods alternating in a simple sinusoidal manner. An important fact brought out by the deep-sea climatic curves is the abruptness of the transition from warm to cold. It is not possible to determine just how abrupt these transitions are because normal deep-sea sediments are gently stirred by bottom life and wafted by currents. However, along the north coast of Venezuela lies a 700-fathom deep basin in which the deeper waters are anaerobic. No burrowers stir the sediment which is laid down in identifiable annual layers. The trench has been stagnant for the past 10,750 years, but during glacial times, the trench was not stagnant and its waters harbored a cool-water fauna. In less than 200 years, the trench stagnated as the surface waters abruptly warmed.

Concern over the meaning of this abrupt transition led to a new theory of glaciation, one which accounts for the alternation from glacial to nonglacial climate in the order of one year's time. M. Ewing and W. L. Donn's theory that the formation or disappearance of the Arctic Ocean's ice cover could profoundly alter world climate is a direct result of this concern over the interpretation of the climatic curves. According to this theory, the

Table 5. Distribution of planktonic Foraminifera in Core A180-73

Sample position in cm from top	*G. m. men.*	*G. m. tum.*	*G. m. flex.*	*G. p. hirs.*	*G. p. punct.*	*G. trc. R.*	*G. trc. L.*	*G. scit.*	*G. infl.*	*G bull.*	*G. pach*	*G. egg.*	*G. rub.*	*G. sac.*	*G. cong.*	*O. uni.*	*Pul. obl.*	*Sph. deh.*
Top	V	A	X	X	C	21	0	A	X	X	X	C	V	V	X	R	V	R
10	V	V	X	R	C	17	0	A	X	X	X	A	V	V	X	A	V	C
20	R	R	X	R	V	101	2	V	X	X	X	A	V	V	X	A	V	X
30	X	R	X	X	V	100	3	F	V	X	X	V	V	V	R	X	X	X
40	X	R	X	X	V	100	3	X	V	X	X	V	V	V	X	R	R	X
50	X	X	X	X	V	100	1	F	V	X	A	V	A	V	R	X	X	X
60	X	X	X	X	V	100	6	R	A	X	V	V	V	V	A	A	X	X
70	X	X	X	F	V	100	5	R	V	X	C	V	V	V	F	C	X	X
80	X	X	X	X	A	100	4	F	V	A	C	V	V	V	F	C	X	X
90	X	X	X	X	V	100	3	R	V	F	C	V	V	V	A	A	X	X
100	X	X	X	R	A	100	3	R	V	X	R	V	V	V	R	X	X	X
110	X	X	X	R	A	100	3	F	V	X	A	V	V	V	R	F	F	X
120	X	X	X	X	A	100	3	R	A	X	A	V	V	V	X	F	A	X
130	X	X	X	X	A	100	1	A	A	X	A	V	V	V	X	R	V	X
140	X	X	X	X	A	100	2	F	A	A	A	V	V	V	X	R	V	X
150	X	X	X	X	A	100	2	F	F	X	C	V	V	V	R	X	V	X
160	X	X	X	R	A	100	2	C	A	X	A	V	V	V	R	A	V	X
170	X	X	X	X	A	100	7	F	C	X	A	V	V	V	R	C	V	X
177	X	X	X	X	A	100	1	R	C	R	V	V	V	V	A	C	V	X
193	X	X	X	X	V	100	4	X	F	X	V	V	A	V	A	R	V	X
200	X	X	X	R	C	100	45	R	F	X	X	V	A	V	R	R	V	X
210	R	X	X	X	C	54	100	C	C	R	R	V	A	V	R	R	V	R
220	F	X	R	R	V	100	8	C	A	X	X	V	V	V	R	R	V	R
230	A	F	F	X	V	100	2	C	A	X	A	V	V	V	R	V	V	R
240	V	V	A	X	V	100	5	V	F	X	R	V	V	V	C	F	V	R
250	V	V	V	X	A	100	4	X	C	R	F	V	V	V	A	A	V	X
260	V	V	V	R	C	140	3	R	A	X	X	A	V	V	C	X	V	X
270	V	V	V	X	F	130	3	F	A	X	X	V	V	V	C	R	V	R
280	A	V	V	X	A	100	4	C	A	F	R	V	V	V	A	V	V	X
290	A	V	V	X	A	100	2	R	R	R	R	V	V	V	V	V	V	X
300	A	A	A	F	C	100	5	X	X	X	X	V	V	V	C	R	V	X
310	V	V	V	R	C	100	6	X	X	X	X	A	V	V	R	R	V	X
320	V	V	V	R	C	100	6	X	X	X	X	A	V	V	C	R	V	R
330	V	V	V	X	C	100	12	R	R	X	X	V	V	V	A	F	V	R
335	V	V	V	X	V	46	100	C	R	X	X	V	V	V	A	F	V	C
340	V	F	F	X	V	44	17	C	R	X	X	A	V	V	C	C	V	C
350	V	F	X	X	A	76	3	F	R	X	X	V	V	R	C	V	V	C
360	A	X	X	X	A	100	3	R	A	X	A	A	V	V	F	F	V	R
370	C	X	R	X	A	100	2	R	V	F	V	A	V	V	R	C	V	X
380	X	X	X	X	A	100	3	R	C	V	A	V	V	V	R	C	X	X
400	R	X	X	F	V	100	2	R	V	C	X	V	V	V	F	C	X	X
440	A	A	X	X	V	100	3	R	A	X	V	V	V	V	F	R	X	X
460	V	X	F	X	C	36	2	C	V	X	C	V	V	V	F	C	F	X
490	V	V	V	R	A	59	3	F	A	R	R	V	V	V	C	X	V	X

SOURCE: *Data from D. B. Ericson and G. Wollin, 1956. X, absent; R, rare, F, frequent; C, common;* A, *abundant;* V, very abundant.

freezing of the Arctic Ocean during one season brought an end to each glacial epoch.

The occurrence of early Tertiary ooze with high carbonate content beneath abyssal red clays indicates a change in the temperature or rate of circulation of the abyssal waters. On the basis of oxygen isotope measurements, the abyssal waters of the Pacific are believed to have cooled 12C° since the Eocene *See* GLACIAL EPOCH; *see also* CLIMATIC CHANGE; PALEOCLIMATOLOGY; POSTGLACIAL VEGETATION AND CLIMATE. [B.C.H.]

Bibliography: W. S. Broecker, M. Ewing, and B. C. Heezen, Evidence of an abrupt change in climate close to 11,000 years ago, *Am. J. Sci.*, vol. 258, 1960; C. Emiliani, Pleistocene temperatures, *J. Geol.*, 63(6):538–578, 1955; D. B. Ericson and G. Wollin, Micropaleontological and isotopic determinations of Pleistocene climates, *Micropaleontology*, 2:257–270, 1956; M. Ewing and W. L. Donn, A theory of ice ages, *Science*, 123(320):1061–1066, 1956.

SAMPLING AND CORING DEVICES

The bottom-sampling devices used to collect sediments from the ocean floor are of three main types: snappers, dredges, and coring tubes. Snap-

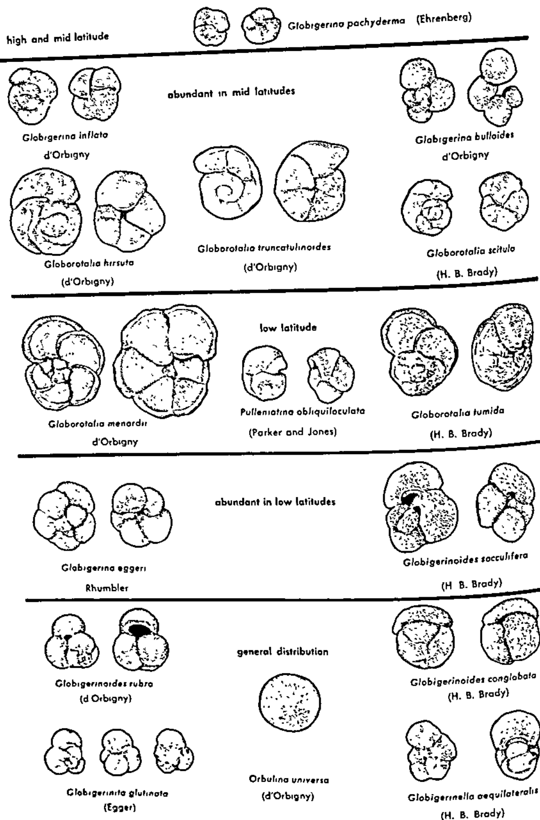


Fig 10. Temperature-sensitive Foraminifera and their latitudinal distribution in the North Atlantic. (After F. B. Phleger, Foraminifera and deep-sea research. Deep-Sea Research, 2(1):1–23, 1954)

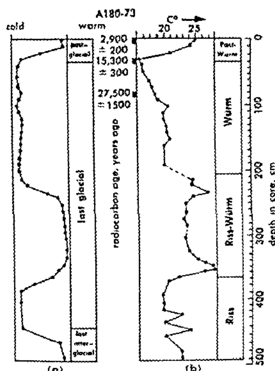


Fig. 11. Climatic fluctuations as determined by analyses of deep-sea sediments. (a) Based on faunal record presented in Table 5 (*after D. B. Ericson and G. Wollin, Deep-Sea Research, 3(2):104–125, 1956*) (b) Based on oxygen isotope measurements (*after C. Emiliani, 1955*).

per-type samplers, with closing jaws actuated by a tension spring and trigger mechanism, are used to obtain small samples from the sediment surface. Snapper, or grab, samplers are generally used in shallow waters where it is desirable to gather a large number of samples as rapidly as possible. The dredge is essentially a bag made of steel links; its mouth is held open by a rectangular frame provided with a bail to which the ship's trawl wire is shackled. When it is dragged along the bottom the dredge collects relatively large objects such as manganese oxide and phosphate nodules, sharks' teeth, and ice-rafted rocks lying on or near the sediment-water interface (Fig. 12).

Coring devices are used to obtain samples of bottom material in situ. This device consists essentially of a steel tube, a glass or plastic liner that may be removed without disturbing the sample, attached weights, and a core-catcher ring and cutting edge which fit the bottom, or penetration edge, of the tube. The amount of sample collected depends on the length of the corer, the weight, and penetrability of the bottom. Core samples up to 12 m in length may be obtained when the smaller types of corer are used. Modifications in coring devices include the piston corer and free-fall re-

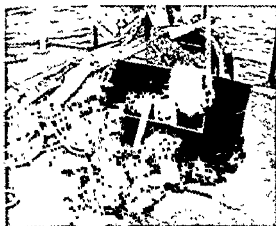


Fig. 12. Bottom dredge showing bottom sample. (*U.S. Navy Hydrographic Office*)

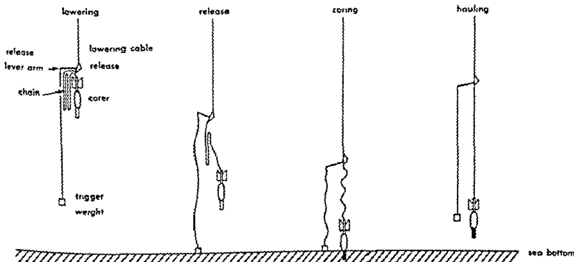


Fig. 13. Principle of operation of free-fall corers. (*From U.S. Navy H.O. Publ. 607, 2d ed., 1955*)

lease mechanism. Both are utilized in the heavier coring equipment used to obtain longer cores.

Piston corer. The piston corer, a steel tube measuring about 60 mm in inside diameter, is capable of recovering undistorted vertical sections of sediment as much as 20 m in length. The corer is driven into the sediment by a free fall of about

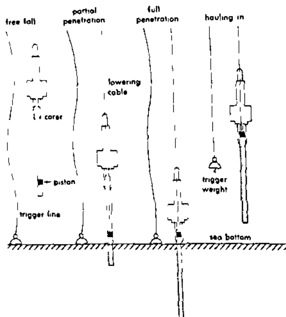


Fig. 14. Principle of operation of piston corers. (From U.S. Navy H.O. Publ. 607, 2d ed, 1955)

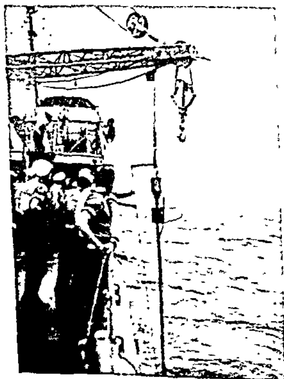


Fig 15. Lowering the Ewing piston corer. (U.S. Navy Hydrographic Office)

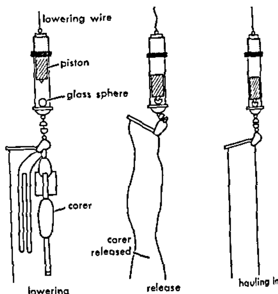


Fig. 16. The ball-breaker principle of operation (From U.S. Navy H.O. Publ. 607, 2d ed., 1955)

5 m and by 500–1000 kg of lead attached to the upper end. Free fall is effected by a trigger weight attached by a wire line to the end of a trigger arm, which rises and releases the coring tube when the trigger weight comes to rest on the sediment surface (Fig. 13). Connection between the ship's cable and the apparatus is maintained by a length of wire rope which passes from the end of the ship's cable down through the coring tube to a piston which is initially at the bottom of the tube. The length is so adjusted that the wire rope becomes taut just when the bottom of the coring tube reaches the sediment surface (Fig. 14). As the tube sinks into the sediment, the piston, immobilized by the cable to the ship, makes hydrostatic pressure outside the tube effective in overcoming frictional resistance between the entering sediment and the inner wall of the tube.

The apparatus is raised by the wire rope attached to the piston which is retained by shoulders near the top of the tube. The bottom of the tube is armed with a removable steel cutting edge. Just above this is the core catcher, a circular set of leaves of spring bronze which close as the tube is withdrawn from the sediment.

The original piston corer of B. Kullenberg (Sweden) is provided with a metal liner in which the core is removed. A simplified corer designed by M. Ewing is used without a liner, and the core must therefore be pushed out of the tube with a plunger and rod (Fig. 15). A pilot core about 30 cm long is usually taken with each long core by means of a small tube with plastic liner attached to the trigger weight.

Bottom contact. In bottom-sampling operations it is necessary to know when bottom contact is made because any excessive wire that is payed out usually kinks. Close attention must be given to sonic depth, wire angle, and length of wire payed out. In shallow waters contact with the bottom may

be detected by a slack in the wire and jerk in the meter wheel. In deeper waters a spring scale or dynamometer may be used to register variations in wire tension. A bottom-signaling device, or ball-breaker, is sometimes used for this purpose. Upon contact with the bottom a small glass ball is crushed (Fig. 16). The resulting implosion signal may be seen on the pen trace of the echo sounder or heard as an audible signal on suitable monitoring equipment. [D.B.E.]

Bibliography: H. Barnes, ***Oceanography and Marine Biology***, 1959; J. B. Hersey, Acoustically monitored bottom coring, *Deep-Sea Research*, 6(2): 170–172, February, 1960; ***Instruction Manual for Oceanographic Observations***, U.S. Navy H.O. Publ. 607, 2d ed., 1955; B. Kullenberg, The piston core sampler, *Svenska Hydrograf. Biol. Komm. Skrifter*, [3]1(2):1–46, 1945.

Marl

An argillaceous, nonindurated calcium carbonate deposit that is commonly gray or blue-gray. It is somewhat friable, and in some respects resembles chalk, with which it is interbedded in some localities. It is formed in some fresh-water lakes, partially by the action of some aquatic plants The plants extract carbon dioxide for photosynthesis from bicarbonate in the water and locally reduce solubility of calcium carbonate next to the leaves, where it precipitates. The precipitate forms a scale, which then falls to the bottom and accumulates. The clay content of marls varies, and all gradations between small amounts of clay (marly limestones) and large amounts (marly clay) are found.

Marlstone, or marlite, is an indurated rock of the same composition as marl. The marlstones are not fissile but blocky and massive with subconchoidal fracture. *See* LIMESTONE; SEDIMENTARY ROCKS.

[R.S.]

Mars

The fourth planet in the solar system. It is visible to the naked eye as a bright red star, except for short periods when it is near its conjunctions with the Sun. It is in opposition with the Sun at intervals which average 780 days. Its orbit has a semimajor axis (mean distance to Sun) of 142×10^6 miles. Its eccentricity of 0.093 causes the distance to the Sun to vary from 126×10^6 mi at perihelion to 155×10^6 at aphelion. Its sidereal period of revolution is 686.98 days; the mean orbital velocity, 15.1 mi/sec; and the inclination of the orbital plane to the ecliptic is 1°51′. *See* PLANET.

The apparent diameter of the disk as viewed from Earth varies from a minimum of 3″5 at conjunction to a maximum of 25″1 at the most favorable perihelic opposition, at which time the distance to the Earth is only 35×10^6 mi; at aphelic oppositions the apparent diameter may not exceed 13″8. Although Venus in conjunction is nearer to the Earth, it is then visible only during daytime and presents its dark side to the observer, while Mars at its nearest is seen at night and appears fully illuminated under the best conditions for observation.

Phases. As an exterior planet, Mars presents only gibbous phases when it is not in opposition or conjunction with the Sun; the maximum phase angle Sun-Mars-Earth is 48° when Mars is in quadrature. The apparent visual magnitude of Mars in conjunction and at its mean distance is +1.5; at mean opposition it is −1.9, at aphelic opposition it is −1.2, and at perihelic opposition it reaches −2.9, at which time Mars appears brighter than any other planet except Venus. Within the small range of the phase angle i observable, the apparent magnitude m, corrected for the effect of varying distance to the Earth and Sun, follows the linear relation

$$m = m_0 + \mu i$$

where m_0 is the magnitude at unit distance and mean opposition and where the phase coefficient μ has the following values: 0.015 degree^{-1} in yellow light, 0.022 in blue and violet light, 0.012 in orange and red light The color-dependence of the phase coefficient indicates that Mars is redder at large phase angle, an effect attributed to its atmosphere.

Albedo. The visual albedo of Mars is about 0.15, about twice the value for the Moon or Mercury, but much less than the values for the planets surrounded by heavy atmospheres. In blue and violet light the albedo is only 0.05; in orange and red light it reaches up to 0.3 (*see* ALBEDO). The high albedo for red light is due to the red color of the surface of Mars.

Elements. The diameter of the globe is 4200 mi ±1%. The relative volume of Mars is therefore about 0.150 (Earth = 1). The polar compression is more in doubt. Dynamical determinations from the rotation of the lines of nodes and apsides of the orbits of the satellites give an ellipticity $(r_e - r_p)/r_e = 1/192$ with an uncertainty of 1 or 2% only. (In this equation r_e is the equatorial radius and r_p is the polar radius.) Direct optical measurements of the disk, however, indicate a higher value, about $1/77$, which exceeds the theoretical upper limit of $1/175$ set by celestial mechanics for a homogeneous planet in hydrostatic equilibrium. The cause of this discrepancy was unknown as of 1959.

The mass of Mars, 0.107 (Earth = 1), is well determined from the motions of its two satellites. The mean density computed from this mass is 3.94 g/cm³ with an uncertainty of a few per cent. The acceleration of gravity at the surface is about 0.38 (Earth = 1), or more precisely, 3.75 m/sec² at the equator and 3.80 m/sec² at the poles, also with an uncertainty of a few per cent. The escape velocity is 3.1 mi/sec.

The rotation period has been very accurately determined from several centuries' observations of the transit times of surface spots at the central meridian of the disk. The sidereal period is 24h 37m 22.67sec and the corresponding length of the mean solar day is 24h 39m 25.0sec. There are, con-

sequently, about 646 Martian sidereal days and 645 mean Martian solar days in the Martian year. The inclination of the equator to the orbital plane is 25.0°, only slightly greater than for the Earth. The seasons are therefore similar but longer. They are, however, more unequal because of the greater orbital eccentricity. The durations of the seasons (in the northern hemisphere of the planet) are as follows: spring, 199 terrestrial days; summer, 182 days; autumn, 146 days; and winter, 160 days. Furthermore, because the perihelion (heliocentric longitude 335°) is close to the summer solstice in the southern hemisphere (heliocentric longitude 357°), the cold season is longer and colder and the warm season hotter and shorter in this hemisphere. This asymmetry has a pronounced effect on the polar [illegible]

disk marked by complex, semipermanent dark regions and variable white polar caps. The same details are observed with greater contrast in orange, red, and infrared light; but in blue, violet, and ultraviolet light an atmospheric veil of clouds or haze obscures the surface details most of the time (Fig. 1). Apart from the atmospheric veil and the large bright and dark regions of the surface, fine details are present—the so-called canals and oases.

ATMOSPHERE

The presence of a tenuous atmosphere around Mars is indicated by the seasonal waning and waxing of the polar caps, by the obscuring blue haze, and by variable cloud formations.

Mass and surface pressure. The total mass and surface pressure of the atmosphere may be estimated by applying the laws of molecular scattering of light to photometric or polarimetric measurements of the light scattered by the atmosphere when free of clouds or haze. Modern determinations are in fair agreement to indicate $P_0 = 85$ millibars = 1.23 lb/in^2 ± 10 to 20% as the most probable value of the atmospheric pressure at the surface. The corresponding mass per unit area is 0.22 (Earth = 1) and the reduced thickness of the atmosphere is 1.10 mi at standard temperature and pressure (STP). The pressure at various altitudes, estimated through the barometric formula for an isothermal atmosphere of molecular nitrogen at an average temperature of 223°K, is given by Fig. 2. Because of the lower acceleration of gravity on Mars, the pressure decreases more slowly with altitude than in the terrestrial atmosphere, and above 25 or 30 km the pressure is actually greater in the Martian atmosphere.

Chemical composition. The chemical composition of the atmosphere derived from spectroscopic observations and theoretical considerations is given in Table 1.

As shown in the table, the main constituent is probably nitrogen, with small amounts of argon and carbon dioxide, the latter being much more abundant than in the atmosphere of the Earth. Argon-40 is believed to be present as a decay prod

Fig. 1. Photographs of Mars taken through color filters at Lick Observatory. Surface details are visible in green, yellow, red, and infrared light; atmospheric details are visible in violet and ultraviolet light

Table 1. Probable composition of the atmosphere of Mars compared with Earth

Gas	Mars		Earth	
	Thickness, m, STP	Volume, %	Thickness, m, STP	Volume, %
Nitrogen, N_2	1650	93.8	6246	78.08
Oxygen, O_2	<2	<0.1	1676	20.94
Argon, Ar	70?	4.0?	74	0.94
Carbon dioxide, CO_2	40*	2.2*	2.2	0.03

* Approximate determinations

uct of radioactive potassium-40 in the crust. As of 1959, all attempts to detect oxygen and water vapor had failed, and only traces of these gases can be present in the atmosphere of Mars.

Vertical structure. The vertical structure of the Martian atmosphere is rather poorly known. Carbon dioxide may play a role similar to water vapor on Earth and lead to the formation of a quasi-stratospheric region, but the altitude of the tropopause (the portion just below the stratosphere) is very uncertain, with values around 6 mi as the most plausible estimates. The vertical gradient of temperature for an atmosphere in convective equilibrium depends only on chemical composition and gravity and is fairly well known—it is about 3.7°C/km. If the mean daytime surface temperature is 270°K, the temperature near the tropopause may be 230–240°K, decreasing more slowly in the quasi-stratosphere perhaps down to 130–140°K above 60 mi. However, absorption of solar ultraviolet radiation by atmospheric gases at high altitudes may greatly affect the temperature distribution. Carbon dioxide is dissociated into carbon monoxide and atomic oxygen by radiations of wavelengths shorter than 1700 A at altitudes of the order of 90 mi, while ionization of nitrogen in an "E layer" similar to the E layer in the Earth's atmosphere may take place near 130 mi. Radiations between 2000 and 3000 A may reach the surface of Mars with appreciable intensity, but this is not definitely known.

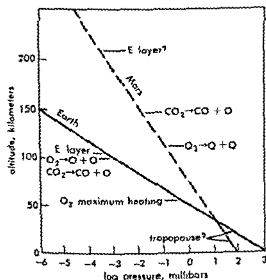


Fig 2. Atmospheric pressure as a function of altitude. (After R. M. Goody)

At wavelengths greater than 3000 A the "blue haze," or "violet layer," scatters and absorbs enough to obscure the surface markings. This may be due in part to ground haze in the lower convection zone, but more probably to an atmospheric mist of ice crystals floating at high altitude. The average diameter of the particles must be about 0.3–0.4 microns (μ) in order to explain the selective absorption. The thickness of the layer may be a few kilometers in order that the observed frequency of occasional clearing of the haze may be explained by plausible temperature fluctuations. However, it is not known why the haze clears most frequently and completely when Mars is near opposition, as has been repeatedly observed since 1926.

Clouds. Three main types of clouds are distinguished, conventionally designated as "white," "blue," and "yellow." The white and blue clouds are mainly observed at the sunrise or sunset limbs, or above the polar regions, especially during fall and winter. The blue clouds, observed in blue and violet light, vanish in red and infrared light; they are therefore tenuous condensations of ice particles generally less than 0.5 μ in diameter. The white clouds frequently observed at the sunrise limbs may be low-lying icy fogs or surface deposits of white frost due to radiative cooling at night; they usually dissipate after 1–2 hours exposure to sunlight. The white clouds observed at the sunset limb are often larger and brighter; occasionally they form day after day above the same localities, as if due to recurrent afternoon convection phenomena. The yellow clouds appear mainly when the planet is near perihelion, they are probably due to storms and enhanced convection raising the finer particles of the surface dust into the atmosphere. The motions of these clouds, which have been followed occasionally for several days, give an indication of the wind velocities and general circulation in the Martian atmosphere. Winds up to 60 mph have been recorded in the early phases of a dust storm, but later in the storm much lower values prevail, progressively decreasing to 6 mph. However, the levels to which such winds refer are uncertain because the altitudes of the various types of clouds are still poorly known, perhaps of the order of 3 mi or less for some yellow clouds and 6 mi for many blue and white clouds. However, the tops of some yellow clouds have been occasionally observed to reach 20 mi or so above the surface. The vertical gradient of the westerly component of the wind is theoretically estimated to be of the order of 1–4 ft/sec per 1000 ft. This is consistent with the observed wind velocities and probable altitudes. For a detailed discussion of the Earth's atmosphere, much of which is applicable to Mars, see ATMOSPHERE.

SURFACE FEATURES

Surface temperature. Both the average and maximum surface temperatures of a planet with tenuous atmosphere can be derived theoretically from Stefan's law of radiation and the known value of the solar constant. The computed temperatures for the Earth and Mars, first assuming black-body radiation, then allowing for the losses measured by the albedo (assumed equal to the visual albedo at all wavelengths), are compared with the observed maximum, average, and minimum temperatures in Table 2. For a discussion of Stefan's law and black-body and gray-body radiation, *see* HEAT RADIATION.

Table 2. Theoretical and observed surface temperatures (°K) of Earth and Mars

	Albedo (visual)	Black body		Gray body		Observed		
		T_{max}	T_{av}	T_{max}	T_{av}	T_{max}	T_{av}	T_{min}
Earth	0 39	394	278	349	246	350	288	200*
Mars	0 15	318	223	307	217	300*	230*	150?

* Approximate determinations

The observed temperatures were determined by radiometric measurements of the infrared radiation of the planet at the Lowell and Mount Wilson Observatories during a period beginning in 1922.

The diurnal variation near the Martian equator is very similar to that observed at the surface of deserts on Earth but is 40–50° lower because of the greater distance to the Sun. The diurnal maximum is reached about 1 hour after noon; the minimum is presumably reached shortly before sunrise. The *total amplitude* of the diurnal variation is about 60°. This is larger than on Earth, as could be expected theoretically in view of the extreme dryness of the atmosphere (see Fig. 3).

The seasonal variations are roughly similar to those of Earth but again have greater amplitude; the maximum temperature may be reached about

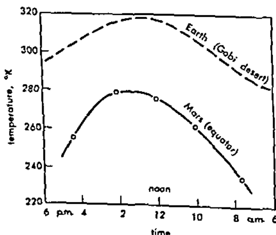


Fig. 3. Diurnal variations of temperature at the surface of Mars. (*After F. Gifford*)

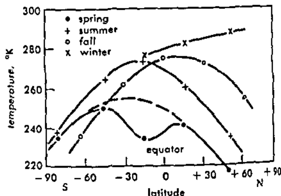


Fig. 4. Seasonal variations of noon surface temperature on Mars. (*After F. Gifford*)

3 weeks past summer solstice (Fig. 4). In addition to seasonal variations, the eccentricity of the orbit causes a difference of about 27° between perihelic and aphelic temperatures of the point directly under the Sun. Furthermore, the dark regions absorb more solar radiation than the bright regions, and are therefore hotter; the maximum difference is about 8–10° for the darkest tropical regions near noon in summer.

Surface temperatures at night can only be estimated from theory. The tenuous, dry atmosphere offers little protection against radiative cooling, and the temperature may drop to 200–220°K at the end of the night, even in the tropical regions. The minimum winter temperature during the long polar night may be as low as 160–170°K. It is unlikely that the temperature drops below 150°K for any length of time since atmospheric carbon dioxide would then condense on the surface, and there has been no evidence of this.

Polar caps and water. The seasonal waxing and waning of the bright polar caps indicates the presence of water on Mars. Spectral, radiometric, and polarimetric observations all prove that the material of the polar caps is ordinary water-ice crystals and not frozen carbon dioxide. Formed during fall and winter by condensation and deposition of the icy mist covering the polar regions, the polar caps extend at the end of winter to latitude 60° in the southern hemisphere and 70° in the northern hemisphere, covering areas of 4×10^6 sq mi and 1.6×10^6 sq mi, respectively. The average thickness of the polar caps is not known accurately, but may be equivalent to about 1 cm of solid ice. The corresponding total mass of water on Mars is 10^{17} g, but the uncertainty is a full order of magnitude. At any moment only a small fraction of this water is in the atmosphere, and most of it is in the form of minute ice crystals forming the blue haze and clouds. The maximum amount of precipitable water vapor is probably less than 10^{-3} g/cm², equivalent to a layer of water 10 μ thick. This is in agreement with the failure of all attempts to detect water-vapor absorption bands and lines in the spectrum of Mars. For all practical purposes

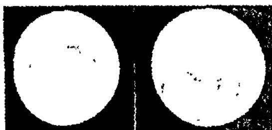


Fig. 5. Telescopic appearance of Mars. (After G de Vaucouleurs)

there is no free water vapor in the Martian atmosphere; thus the possibility of common meteorological phenomena such as rain, hail, snow, or clouds of liquid droplets is excluded.

The regression of the polar caps in spring and summer is therefore brought about by superficial sublimation rather than by ordinary "melting."

Bright regions. The reddish areas which cover about three-quarters of the surface of Mars are vast desert expanses covered by a very fine dry dust, colored red probably by iron oxides. Although some moderate relief is probably present, high mountain chains have not been observed.

Faint, irregular streaks or patches, the so-called canals and oases, are observed in the bright regions (Fig. 5), but their exact structure is difficult to define and their real nature is unknown. However, the existence of the extensive network of long narrow lines drawn by many observers in the bright and dark regions at the turn of the century has not been confirmed by the best modern visual and photographic observations.

Dark regions. The dark areas of the surface of Mars form a conspicuous and fairly stable pattern which has been mapped in detail (Fig. 6). The dark regions are, however, subject to seasonal variations in albedo, polarization, and sometimes shape which are closely related to the concomitant cycle of the polar caps (Fig. 7). These variations are usually regarded as associated with the atmospheric transfer of water vapor from pole to pole

Fig. 6. Map of Mars by G. de Vaucouleurs (based on observations of 1939 and 1941).

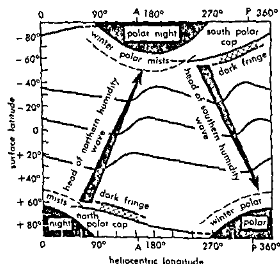


Fig. 7. Seasonal cycle of polar caps and dark regions of Mars. The latitude of the edge of the polar caps and the variations of intensity of the dark regions at various latitudes are schematically shown. The cross-hatched portions indicate the period of visibility of a localized darkening of great intensity. A, aphelion; P, perihelion

and are often taken as evidence for the existence of plant life on Mars (Fig. 8).

Another important property of the dark regions is the fairly frequent occurrence of irregular, nonseasonal variations. On several occasions dark regions have encroached over previously bright territory for periods of from a few months up to several decades. In general, the region affected returns sooner or later to its previous "normal" aspect. Some regions are more liable than others to such instability. Some of the canals and oases share in this activity, while others seem much more stable. Occasional, short-lived dark spots have also been seen to appear in isolated places of the bright areas. The cause of such changes is unknown but they indicate a great deal of activity on the surface.

Two main hypotheses have been advanced as to the nature of the dark regions. (1) The volcanic hypothesis attempts to account for the structure, shape, and variations of the dark regions by the assumed presence of suitably located volcanoes acting as sources of vast volumes of ash and cinders which are carried by the prevailing winds and deposited over certain areas. The longer and hotter warm season in the southern hemisphere is taken as the cause of the preponderance of deposition in this hemisphere. (2) The vegetative hypothesis assumes the presence of a low form of vegetation, for example, lichens, microscopic algae, or bacteria, in the dark regions. In addition to the seasonal and irregular variations of the dark regions and their persistence in the presence of secular aeolian deposition of desert dust, this hypothesis is supported by another fact. This is the discovery in the infrared spectrum of the dark regions of Mars of absorption bands of the C—H bond which are characteristic of organic molecules and which are observed also in the reflection spectrum of terrestrial vegetation. This observation, originated by W. M. Sinton in 1956 and confirmed by him in 1958, comes close to providing the final proof of the existence of plant life on Mars.

INTERNAL CONSTITUTION

The internal constitution of the globe cannot be uniquely determined, but the plausibility of various hypotheses can be tested by comparing the predicted and observed polar flattening. The most plausible hypothesis assumes a small nickel-iron core of density about 10 g/cm³ and radius 700 km, surrounded by a magnesium silicate (olivine) mantle of density varying from about 3.8 g/cm³ at the limit of the crust to about 4.3 g/cm³ at the limit of the core. The crust, about 200 km thick, may have a density varying from 3.28 g/cm³ at the surface to 3.36 g/cm³ at the limit of the mantle. These figures are subject to considerable uncertainty.

SATELLITES

Mars has two small satellites, Phobos and Deimos, discovered by A. Hall in 1877. Although they are not intrinsically very faint, a large telescope is required to see them clearly against the bright glare surrounding the planet.

The outer satellite, Deimos, moves at a mean distance from the planet's center of 14,600 mi in a sidereal period of 30h 18m, only slightly longer than the period of rotation of the planet. As a result, the apparent period with respect to a point on the surface of Mars is about 132 hours, over four times the true period of revolution. The orbital plane is inclined about 1° to the equatorial plane of the planet.

The inner satellite, Phobos, moves at a mean distance from the planet's center of 5850 mi in a sidereal period of 7h 39m, which is less than one-third of the mean Martian solar day. It is the only known satellite whose period of revolution is less than the period of rotation of the parent planet; because of this Phobos, as seen from the surface of Mars, rises in the west and sets in the east, moving in a direction opposite to that of all other known celestial bodies. As a result, the apparent period with respect to a point on the surface of Mars is about 11 hours and therefore Phobos rises and sets twice each Martian day. The inclination of the orbital plane to the equatorial plane of the planet is 1.7°. The mean distance of Phobos to the surface of Mars is only 3750 mi and the apparent diameter of Mars seen from Phobos is 42°; conversely, Phobos cannot be seen from the polar regions of Mars within 21° from the poles.

The diameters of the Martian satellites are too small to be measured directly. The true linear diameters can be derived only from the apparent visual magnitudes at mean opposition, about 11.5 and 12.5 for Phobos and Deimos respectively, and the plausible assumption that the albedo of their

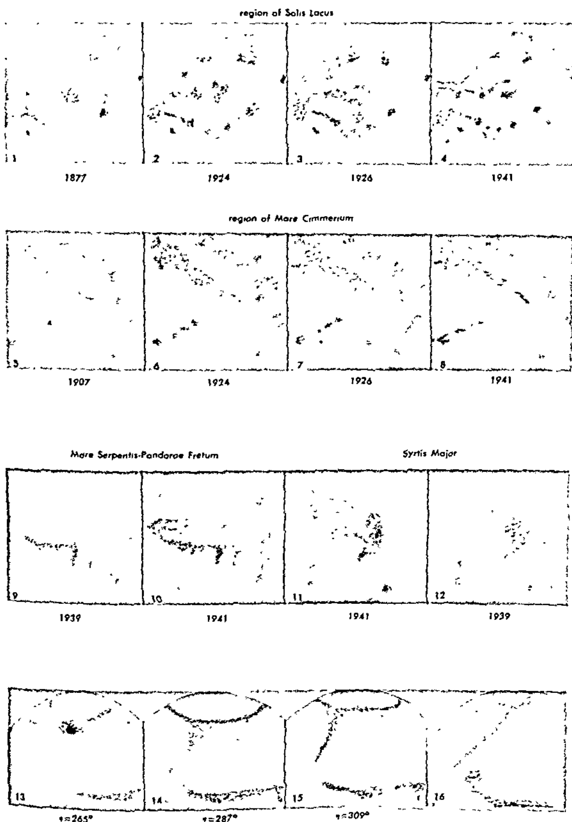


Fig. 8. Irregular variations (top two panels) and seasonal variations (bottom two panels) on the surface of Mars. The symbol η in the bottom panel refers to the heliocentric longitude of Mars.

surfaces is the same as that of Mars, namely 0.15. The diameters so computed, about 10 mi for Phobos and 5 mi for Deimos, are probably correct within a factor of 2.

The attraction of the equatorial bulge of Mars causes the lines of the apsides of the satellites' orbits to advance and the lines of the nodes to retrograde in a period of about 55 years for Deimos and 2.2 years for Phobos. From this motion the dynamical flattening of Mars can be accurately derived.

CANALS OF MARS

One of the most peculiar features of the Martian surface is the presence of faint, narrow, linear markings which have given rise to much speculation and discussion since their discovery by the Italian astronomer G. V. Schiaparelli in 1877. Schiaparelli, and later P. Lowell in the United States, G. Fournier in France, and others, represented Mars as covered by a network of hundreds of exceedingly narrow lines intersecting in small, round, dark spots or oases.

These lines, first observed in the bright regions only, were later seen to cross the dark regions also. Most canals followed great circles of the sphere; no canal was seen to vanish in the bright desert regions, but always would begin and end in an oasis, another canal, or in an "estuary" on the shore line of a dark region. Furthermore, some of the originally single lines were at times replaced by two nearby parallel lines; this was the strange phenomenon of the gemination of the canals. Finally, the visibility or strength of the canals seemed to vary with the seasons or, more exactly, in relation to the phase in the annual cycle of each polar cap and according to latitude, that is, to vary in synchronism with the general seasonal wave of darkening of the dark regions.

Eventually, all the fine shadings of the Martian surface were somehow overlooked by Lowell and his school, and maps of the planet published at the turn of the century took an appearance similar to that of a spider web or a railroad map. This inevitably led to the speculation that the canal network was artificial and gave "proof" of the existence on Mars of intelligent inhabitants with an advanced technology. According to Lowell's interpretation of his observations, the canal network represented a monumental, planet-wide organization for the capture, distribution, and utilization of the small amount of water available in the polar caps. Each canal was regarded not as the water-carrying ditch or pipeline, but as the band of vegetation growing on both sides of it in the irrigated area, that is, much as the fertile Nile Valley would signal the presence of the narrower Nile River to an extraterrestrial observer. With the coming of spring the polar canals would first become filled and vegetation would begin to grow around them; then, the flow of water would reach the canal network of the temperate regions, and later that of the equatorial regions. The growth of vegetation would follow, thus accounting for the seasonal variations of visibility and their latitude dependence. The gemination of the canals was explained in the same scheme by the opening up or activation of a second irrigation channel when and where an unusually large amount of water became available. Finally, Lowell interpreted the oases at the meeting places of several canals as primary centers of activity, perhaps the locations of major pumping stations.

When first developed, this theory was both internally consistent and in general agreement with the meager knowledge of physical conditions on Mars available at the time. However, it did not survive the test of subsequent observations and the progress of physical investigations. In the first place, the width of the geometric canals drawn by Schiaparelli and Lowell and the separation between the components of many geminated canals were often much smaller than the resolving power of their telescopes would allow. Then, experienced observers such as E. E. Barnard in the United States, E. M. Antoniadi in France, and others, failed to detect the canal network with more powerful instruments through which they were able to see distinctly much fine, irregular detail of the surface that had eluded the smaller telescopes of Schiaparelli and Lowell. Some unusually sharp photographs occasionally obtained at various observatories, especially since 1941 at the Pic du Midi in France, fully support these conclusions; whenever very fine details are recorded, the extensive canal network of Schiaparelli and Lowell fails to appear. The conclusion seems inescapable that it must have been caused by optical illusion. Finally, present knowledge of physical conditions on Mars, especially of surface temperatures, seems to preclude the possibility of extensive circulation of water, particularly in an irrigation network as envisioned by Lowell.

It should be added, nevertheless, that the presence of minute details in the bright and dark regions of Mars is not contested by experienced observers; furthermore, it is generally conceded that such details show a definite tendency, at least in some regions, to be aligned along preferred directions. For instance, on several occasions observers using large telescopes under favorable conditions have resolved some of the more prominent canals into strings of irregular spots, so that, although the canal does not exist as a narrow, continuous line, it still marks a direction or axis on the surface along which details are clustered. Finally, along such axes and often precisely along the same course previously marked only by an insignificant and barely seen canal, there may occasionally appear conspicuous, broad dark bands lasting for months or even years at a time; a classical example is the Nepenthes-Thoth (see Fig. 6) which alternates frequently between phases of great width and intensity, such as in 1888, 1926, 1958, and longer periods of relative faintness. Other examples are the now vanished Hydaspes, which was strong and broad in the 1860s, and the Nilosyrtis,

the largest and darkest canal on Mars throughout the latter half of the nineteenth century but now visible only with difficulty as a faint, shadowy band resolvable into small, irregular spots with high telescopic powers.

From this and other evidence it can be concluded that, while the geometric canal network is illusory, there exists on the surface of Mars a definite, though variable pattern of small, faint markings whose various parts seem to be associated with or related to some preferential axes of activity. Both the nature and mechanism of such relations and activity remain unknown. [C.D.V.]

Bibliography: G. de Vaucouleurs, *Physics of the Planet Mars*, 1955; G. de Vaucouleurs, *The Planet Mars*, 1950; F. L. Whipple, *Earth, Moon, and Planets*, 1941.

Marsupialia

The only known order of the infraclass Metatheria. Marsupials are characterized by the presence of a pouch (marsupium) in the female in which the young, born in a very immature state, are sheltered for a time after birth. In a few species the pouch is vestigial or has disappeared completely.

Except for the opossum, which is a relatively recent immigrant into North America, living marsupials are confined to the Australian region and to South and Central America. Fossils show that marsupials were much more widely distributed during the Late Cretaceous and early part of the Tertiary. Australia and South America were isolated throughout most or all of the Tertiary, and thus formed refuge areas where these animals were not disturbed by the more progressive placentals that evolved later, and which exterminated marsupials everywhere else. When the Panama land bridge reconnected South America with Central America at the close of the Tertiary, a host of modern placentals streamed southward and quickly exterminated all but a handful of the rich South American marsupial fauna. The survivors are small and primitive forms.

Australia was never reconnected with the Asiatic mainland, and except for monotremes, bats, a few primitive rodents, and the dingo (which was introduced by early man), the Australian therian mammalian fauna is essentially a well-balanced marsupial assemblage. About 50 genera are represented. These have occupied most of the ecological niches, and afford a remarkable parallel with the adaptive radiation of placental mammals elsewhere. Marsupials include a "wolf," "cats," several kinds of "rodents," and even a "mole." The kangaroos and wallabies are the ecological counterparts of deer and antelopes. *See* METATHERIA; ZOOGEOGRAPHY. [D.D.D.]

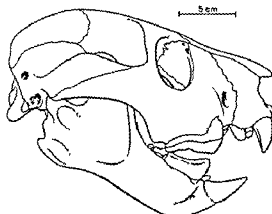
Marsupialia fossils

The marsupials, or pouched mammals, include the opossums, borhyaenids, dasyures, koalas, kangaroos, diprotodontids and many other groups. In their long evolutionary history, inadequately known from the fossil record, marsupials diversified almost as much as placental mammals. Indeed the genetic magnitude of the order is greater than any other order of mammals.

Marsupials are born in the embryonic state and migrate to the marsupium where they attach themselves to imperfectly developed nipples (the gestation period in the opossum is 12–13 days, in the kangaroo, about 40). Epipubic bones are usually present; they support the pouch. The third premolar is the only replacement tooth, and the angle of the mandible is inflected. The alisphenoid bone is expanded into a bullalike structure over the ear bones in many groups. The basic dental formula (as seen in the American opossum, family Didelphidae) is incisors $\frac{5}{4}$, canines $\frac{1}{1}$, premolars $\frac{3}{3}$, and molars $\frac{4}{4}$. The number of teeth is reduced in other genera

The oldest record of marsupials is from the Cretaceous in North America. Even at that time they had become adaptively diversified although they were primarily insectivorous, carnivorous, or omnivorous. Their ancestry has not been traced back into the known Jurassic orders. Opossumlike genera have been recorded for each of the Cenozoic Epochs in North America. In Europe similar forms lived from Paleocene through late Miocene. No marsupials have been discovered in Asia or Africa.

In South America the families Didelphidae, Borhyaenidae, Caenolestidae and Polydolopidae are well represented. Since no placental carnivores or insectivores were present on that continent during most of the Cenozoic Era, the marsupials were able to assume these roles. The most spectacular of the South American marsupials are the hyaenoid *Borhyaena*, the doglike *Lycopsis*, and the sabertoothed *Thylacosmilus*. The remarkable similarity of *Lycopsis* to *Thylacinus* of Australia and Tasmania is one of the best examples of convergent evolution. Caenolestids and polydolopids also display an evolutionary convergence with some of the smaller kangaroos, especially in the construction



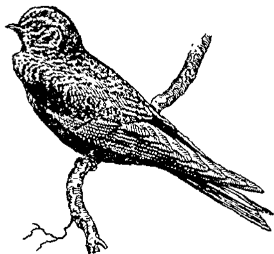
Skull and jaw of *Thylacoleo carnifex*, an aberrant phalangeroid marsupial from the late Cenozoic of Australia. (*After Woods, 1956*)

of their serrate, crested premolars. See CARNIVORA FOSSILS; INSECTIVORA FOSSILS.

It is not known when marsupials reached Australia or how they got there, but evidence from the late Cenozoic faunas indicates that the eight or more families have experienced an evolutionary history that probably extends back into Cretaceous time. Fossils of macropodids, a bushtaillike opossum, a dasyure, and a koalalike animal have been found in Australian mid-Tertiary rocks. In addition to almost all of the living genera, the Australian Pleistocene fauna includes numerous large or gigantic forms, such as *Diprotodon*, which is larger than a rhinoceros; *Phascolonus*, or giant wombat; *Thylacoleo*, the so-called marsupial lion; and numerous giant kangaroos. [R.A.ST.]

Martin

Any one of three species of American swallows of the genus *Progne*, two of which occur in the United States. The purple martin, *P. subis*, nests throughout most of the United States and southern Canada except the Pacific Coast, but it is more abundant in the southern and central states where it nests in special boxes erected for its use. The martin is a



Purple martin, *Progne subis*; length to 8½ in. (From *E. L. Palmer, Fieldbook of Natural History*, McGraw-Hill, 1949)

sleek, long-winged swallow; the male is glossy purple and the female bluish to brownish gray, with dull gray underparts. These birds formerly nested in holes in trees, but were provided with gourd houses by the Indians before the coming of the white man. They winter from Florida to South America. The white-bellied martin, *P. chalybea*, is a tropical American species, reaching the United States only in the lower Rio Grande Valley. See PASSERIFORMES; SWALLOW. [J.D.B.]

Maser

A device for coherent amplification or generation of electromagnetic waves by use of excitation energy in resonant atomic or molecular systems. The word is an acronym for *m*icrowave *a*mplification by *s*timulated *e*mission of *r*adiation. The device uses an unstable ensemble of atomic or molecular particles which may be stimulated by an electromagnetic wave to radiate excess energy at the same frequency and phase as the stimulating wave, thus providing coherent amplification. Masers, however, are not limited to the microwave region; this type of amplification has been extended to include a frequency range from audio to infrared or optical frequencies. Maser-type amplifiers and oscillators are also sometimes referred to as molecular, or quantum-mechanical, since they involve processes on a molecular scale, and since some types cannot be adequately described by classical mechanics, but show characteristic quantum-mechanical phenomena.

Maser amplifiers can have exceptionally low noise, and come close to effectively amplifying a single quantum of radiation in the microwave region; that is, they approach the limits, set by the uncertainty principle, on the precision with which phase and energy of a wave may be amplified. Their inherently low noise makes maser oscillators that use very narrow atomic or molecular resonances extremely monochromatic, providing a basis for frequency standards. See ATOMIC CLOCK. Since atoms or molecules may have resonances and effective amplification over a wide frequency range and to very short wavelengths, masers are useful as coherent amplifiers of millimeter, infrared, or perhaps even optical wavelengths, where older types of circuit elements are not effective.

Because of their low noise and high sensitivity, maser amplifiers are particularly useful for reception and detection of very weak signals in radio astronomy, microwave radiometry, long-distance radar, and long-distance microwave communications. They also provide research tools for very sensitive amplification or detection of electromagnetic radiation.

Thermodynamic equilibrium of an ensemble of particles—such as atoms, molecules, electrons, or nuclei—which have discrete energy levels and which may radiate electromagnetic energy, requires that the number n_1 of particles in a lower level 1 be related to the number n_2 in an upper level 2 by the Boltzmann distribution condition $n_1/n_2 = e^{(E_2-E_1)/kT}$, where E_1 and E_2 are the respective energies of the two levels; k is Boltzmann's constant; and T is the absolute temperature, a positive number. See BOLTZMANN STATISTICS. Thermodynamic equilibrium also requires that phases of oscillation of the particles, or relative phases of quantum-mechanical wave functions for the various states, be random. A violation of either condition can result in instabilities which may release electromagnetic radiation. The frequency ν of radiation released is characteristically given by $h\nu = E_2 - E_1$, where h is Planck's constant.

The particles may be stimulated by an electromagnetic wave to make transitions from the lower

to the upper level, absorbing energy from the wave, or from the upper to the lower level, imparting energy to the wave, and thereby increasing the wave amplitude coherently. Stimulated transitions from the upper to the lower state and those from the lower to the upper state are equally probable. For equilibrium at any positive temperature, the Boltzmann distribution requires that n_1 be greater than n_2. Therefore, there is a net absorption of energy from the wave, because particles which absorb are more numerous than those which emit. If the condition $n_2 > n_1$ occurs, the system may be said to have a negative absolute temperature, because the Boltzmann condition is fitted only by a negative value of T. If there are not too many counterbalancing losses from other sources, this condition allows a net amplification, because particles which emit energy are more numerous than those which absorb.

Gas masers. An amplifier where $n_2 > n_1$ is the beam-type maser (Fig. 1). Operated in 1954, this was the first type of maser to be suggested. Ammonia gas issues from a small orifice into a vacuum system to form a molecular beam. Molecules in the lower of the two states are deflected away from the axis of the state sorter or focuser by inhomogeneous electric fields which act on their dipole moments. Those molecules in the upper state are deflected towards the axis and sent into the microwave-resonant cavity. If losses in the cavity walls and coupling holes are sufficiently small, or if the number of molecules is sufficiently large, amplification or oscillation will occur. This maser is particularly useful as a frequency or time standard because of the relative sharpness and invariance of resonances of the ammonia beam.

The condition for oscillations to occur is

$$n_2 - n_1 \geqq hV\Delta\nu/8\pi\mu^2 Q$$

where h is Planck's constant, V the cavity volume, $\Delta\nu$ the width at half maximum of resonant response of the molecules, μ the molecular dipole moment (matrix element), and Q the quality factor of the loaded cavity. The maximum power

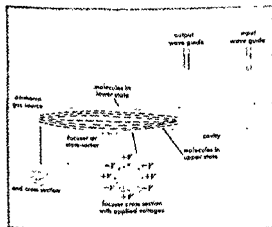


Fig. 1. Schematic of beam-type maser.

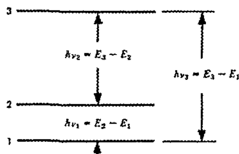


Fig. 2 Energy levels for a three-level maser.

output is approximately $h\nu$ multiplied by the rate at which molecules enter the cavity, and is very small. A wave impinging on the cavity will be amplified on reflection if it occurs near the resonant response and if $n_2 - n_1 \geqq hV\Delta\nu/8\pi\mu^2 Q_0$, where Q_0 is the quality factor of the unloaded cavity.

Other masers using gaseous molecules have been proposed which involve production of a non-equilibrium distribution by excitation of the gas by means of externally-applied radiation of shorter wavelength. Normally, such a system requires molecules with at least three energy levels, two used for amplifying a wave and a third higher level to which molecules are excited from the lowest level, as indicated in Fig. 2. In decaying from the higher level, molecules return, at least in part, to fill up the intermediate level and to satisfy the condition $n_2 > n_1$. If it is light radiation which excites molecules to higher levels, the system is said to use "optical pumping."

Solid-state masers. Solid-state masers usually involve the electrons of paramagnetic atoms or molecules in a static or slowly varying magnetic field. In the simplest case, the two-level solid-state maser, only one electron on each molecule is affected. The energy of the electron is quantized into two levels, according to whether the magnetic moment, associated with the electron spin, is parallel or antiparallel to the magnetic field. At thermal equilibrium there are more magnetic moments parallel than antiparallel to the field, corresponding to $n_1 > n_2$. This situation may be reversed, so that $n_2 > n_1$, by interchanging the two populations n_1 and n_2. The interchange is accomplished by rapid variation of the frequency of an intense electromagnetic field through resonance, by application of a pulse of resonant electromagnetic radiation, or in principle by sudden reversal of the magnetic field. *See* ELECTRON SPIN.

Electron-spin moments are more weakly coupled to the electromagnetic field than are molecular electric-dipole moments (by a factor of about 10^4). A much larger preponderance in the upper state is required than for the maser of Fig. 1. If requirements for amplification are met, however, electron-spin moments give correspondingly greater power output. Furthermore, their resonant frequencies are easily tunable by variation of the magneti

field, because their energies involve interaction between electronic magnetic moments and the field. In the simplest cases, the resonant frequency in megacycles is approximately 2.8 times the magnetic field strength in oersteds. Electron-spin resonances in paramagnetic materials allow amplification over broader bandwidths (one to a few hundred Mc) than do gas systems. Favorable conditions are usually obtainable only at very low temperatures, such as occur in a liquid-helium cryostat; hence, cryogenic problems are often involved and materials used are normally solids rather than liquids. The two-level solid-state maser is most easily operated in pulses, between which the populations of the two levels are readjusted.

The popular three-level solid-state maser also uses paramagnetic material containing electronic magnetic moments in a magnetic field. It has many of the characteristics of the two-level solid-state maser, but can be operated continuously with much more convenience. In Fig. 2, the spacing between the three levels is shown to correspond to microwave frequencies ν_1, ν_2 and ν_3. Usually a few milliwatts of microwave power at frequency ν_3, called [illegible]

plest assumptions, the number of systems in levels 1 and 3 is equal and the number in level 2 is greater if $\nu_1 < \nu_2$, giving amplification at frequency ν_1. If $\nu_1 > \nu_2$, simple assumptions predict amplification for frequency ν_2.

For three suitable levels to occur in a paramagnetic material, each paramagnetic center must involve the magnetic moment of more than one electron, and must interact with a surrounding array of atoms which are not cubically arranged. The energy levels and frequencies still respond to an externally applied magnetic field. They are, however, no longer simply related to it but may vary widely in accordance with fields internal to the crystalline material. This allows responses at high frequencies with relatively low applied fields.

Three-level solid-state maser amplifiers have been made which have a noise temperature less than about 5°K (noise figure ≤ 1.02). Although a wide variety of paramagnetic materials may be used, synthetic ruby, containing paramagnetic chromic (Cr^{+++}) ions, is favored. It has provided amplification both in resonant cavities and in traveling-wave structures.

Maser oscillators in the megacycle and audio range have been proposed; they use nuclear moments in an applied magnetic field or, with pumping at a higher frequency, in internal crystalline fields. The magnetic moments of protons in liquid water have provided a successful maser of this type. Small impurities of a paramagnetic ion furnish the higher energy level needed and transfer their excitation to the protons. The proton resonances in a magnetic field must be extremely narrow. Such a maser may be used as a very monochromatic oscillator with frequency proportional to the magnetic field strength.

Deviation from thermodynamic equilibrium of the second type, involving phase coherence, also allows maser-type amplification and is present in many masers. In the beam-type maser oscillator, molecules decay towards the lower state. They continue to amplify after the probability of their being found in the lower state is greater than that of being found in the upper, because they oscillate coherently and in such a phase that they transfer energy to the electromagnetic field.

Raman-type masers rely on a type of phase coherence. Molecules with two levels, separated by energy $h\nu_1$, may be strongly driven by an electromagnetic field of frequency ν_2. If the majority of systems is in the lower state, and if $\nu_2 > \nu_1$, the Raman effect can allow amplification at frequency $\nu_2 - \nu_1$ (see RAMAN EFFECT). This requires an intense driving field or a very strong coupling of the systems to the field, such as occurs in ferromagnetic electron resonances. In ferromagnetic materials, large numbers of electrons act in unison, thus providing coupling to an oscillating field which is strong enough to make Raman effects prominent. This can also be discussed in terms of classical theory if nonlinearities are allowed for, and is closel[illegible] and to other [illegible] RESONANCE; [illegible] LAR BEAMS; MOLECULAR STRUCTURE AND SPECTRA; NUCLEAR MOMENTS; PARAMETRIC AMPLIFIER.

Maser circuits. Maser circuits characteristically involve atoms or molecules which provide resonant reactances, positive or negative resistances, and coupling between two or more frequencies or circuit components. They also use certain elements of spectroscopic systems and a wide variety of components found in other radio-frequency and microwave devices.

In the simplest cases the molecular resonances behave like a resonant LC circuit with a positive or negative series resistor, or like a large number of such circuits in parallel and tuned over a distribution of frequencies.

The more classical circuit elements normally involved in masers supply the following functions:

1. Means for ensuring sufficiently strong interaction of an electromagnetic wave with material which amplifies the wave by stimulated emission.
2. Input and output coupling for the wave.
3. Auxiliary circuits which take appropriate advantage of maser characteristics in an over-all system.
4. Where electromagnetic excitation is used, circuits which supply energy to the material and produce an unstable state which can radiate.
5. Magnetic field or other components for controlling frequencies of resonance of the material. The schematic of a three-level solid-state maser amplifier, shown in Fig. 3, illustrates each of these functional parts.

The resonant cavity, fulfilling function 1, must have sufficiently low internal losses and produce a sufficiently intense oscillating field (in this case a magnetic field) in the region of the amplifying material. If the cavity is uniformly filled with this material, the condition that amplification be obtainable is

$$Q_0 \gtrsim -Q_M = \frac{h\Delta\nu V}{8\pi\mu^2(n_1 - n_2)}$$

where Q_0 is the unloaded quality factor of the cavity, h is Planck's constant, $\Delta\nu$ the width of the molecular or atomic resonance, V the cavity volume, n_1 the number of particles in the upper state, n_2 the number in the lower state, and μ the effective dipole moment (matrix element) of the atoms or molecules.

A maser of this type is very similar to any other amplifier with positive feedback. There is an effective negative resistance with which may be associated the negative quantity Q_M, giving a fractional gain per cycle in energy stored of $2\pi/|Q_M|$. As the losses, characterized by the loaded Q of the cavity, are decreased, the amplifier gain increases until it becomes unstable and oscillates when $1/Q + 1/Q_M \leqq 0$. Here $1/Q = 1/Q_0 + 1/Q_E$, where Q_E applies to the external coupling. The power gain is $G = [(2Q/Q_E) - 1]^2$. The bandwidth B decreases with gain in such a way that the bandwidth-voltage gain product is nearly constant under typical conditions, or

$$(\sqrt{G} + 1)\ B = 2\nu/Q_E$$

where ν is the frequency for maximum gain. Thus, for given characteristics of the maser material specified by Q_M, the gain and bandwidth can be adjusted over certain limits by variation of the cavity losses, or Q_0, and the coupling, or Q_E.

The noise temperature T for the amplifier alone (Fig. 3) is given by

$$T = \frac{h\nu}{k \ln\left[\frac{n_2}{n_1} + \frac{n_1 - n_2}{n_1}\frac{Q_M}{Q_0}\right]}$$

where k is Boltzmann's constant. Minimum noise temperatures require a Q_0 appreciably larger than Q_M and a small ratio n_2/n_1. A noise temperature of $h\nu/k$, or near 1°K for microwave frequencies, is the minimum needed, since this allows effective amplification of approximately one quantum, the limit set by quantum mechanics.

The input and output use the same coupling hole in Fig. 3, and require a directional coupler, or preferably a circulator, for their separation. Separate input and output coupling holes may be used, but this tends to decrease the gain-bandwidth product and to increase noise.

To take full advantage of the low noise in the maser amplifier, the input and output circuits must be prevented from radiating excess noise into the amplifier. For example, attenuation in parts of the input wave guide and in the circulators, which are not at low temperature, results in noise radiation into the amplifier. If there is 0.1 db loss in some part which is at room temperature, noise radiation into the amplifier corresponding to about 7°K will result. Since most solid-state maser amplifiers operate at low temperatures, some components of the input or output circuits may conveniently be cooled to minimize their noise radiation. If input and output coupling holes are used, the output wave may be passed through a cooled isolator to avoid noise radiation from warm components of the output circuit.

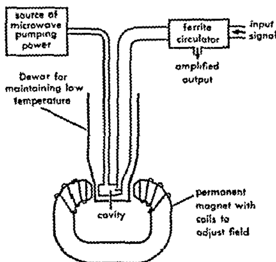


Fig. 3. Block diagram of maser amplifier system for 9000 Mc.

If energy is supplied to the amplifying material by an electromagnetic drive, the circuits used for this purpose must provide a sufficiently strong controllable drive (usually constant in amplitude and frequency). If the interaction between the driving or pumping radiation and the material is strong, a relatively small amount of power may be needed. In Fig. 3 the cavity must be resonant at both the pumping and the amplifying frequencies. In some cases the orientation of fields at both frequencies must also be carefully controlled. Usually excess pumping power is available so that careful design for coupling it to the amplifying material is not essential.

Masers using a resonant cavity may also be operated as super-regenerative amplifiers, for which various components such as the driving field, the cavity losses, the couplings, or the "static" magnetic field may be modulated.

Traveling-wave masers normally use slow-wave structures, which need only be effective over the range of response of the maser material. Amplification of a reflected wave may be controlled in the usual ways by matching, by attenuators or isolators, or by arrangement of the maser material itself to absorb a reflected wave.

The overloading of a maser amplifier is qualitatively like that of any other amplifier, the overload

occurring when the molecular energy becomes exhausted. Recovery times vary between microseconds and seconds, depending especially on the maser material used; *TR* (transmit-receive) devices are sometimes necessary (*see* RADAR).

Maser oscillators used as frequency standards require especially stable cavities, in which the field is strongly coupled to the excited molecules. The cavities should be decoupled from external circuits, as by an attenuator or isolator.

Amplification or generation of electromagnetic waves at infrared or optical frequencies by maser techniques appears to be an attractive possibility. Probably multimode cavities, lenses, directional selectivity of the emitted radiation, and other infrared or optical techniques will be important for these maser applications. [C.H.TO.]

Bibliography: J. R. Singer, *Masers*, 1959; J. P. Wittke, *Proc. IRE*, 45:291, 1957.

Masking (sound)

Interference with the audibility of a sound caused by the presence of another sound. More specifically, the number of decibels (db) by which the intensity level of a sound must be raised above its threshold of audibility (the level at which it is just audible) in quiet, to be heard in the presence of a second sound, is called the masking produced by the second sound on the first. Masking is an important consideration in psychoacoustics because the listener is seldom in a noise-free environment.

Masking between pure tones. Masking effects when both the masking and the masked sounds are pure tones and both are introduced into one ear of a listener are shown in Fig. 1. The frequency of the masking tone is given by the number above each set of curves; its level, in db above its own threshold of audibility, is the number attached to each curve. As the masked tone is moved through the range of frequencies shown on the abscissa and is raised in level until just heard, the amount of masking (or threshold shift) is plotted in the curves. Three major effects are to be noted. First, as would be expected, the higher the level of the masking tone, the greater the masking. Second, the masking is greatest when the frequency of the masked tone is in the vicinity of that of the masking tone. The dips in the curves when the frequencies are very nearly the same are caused by audible beats between the two tones; these make the presence of the masked tone more apparent. Third, the masking caused by a tone is much greater on frequencies higher than that of the masking tone than on frequencies below it. This means that a low-frequency noise will have much greater effect in masking sounds of medium frequency than will a high-pitched noise of the same intensity level.

Subjective tones. The question of what is heard when both tones are audible was explored by R. L. Wegel and C. E. Lane in the first well-controlled tests of masking (1924). In addition to the two tones themselves, other tones are heard which are subjective, since they do not exist except in the listener's hearing. They include harmonics of both tones, tones whose frequencies are the difference and sum of those of the two tones, and sum and difference tones of various harmonics. These subjective tones are presumably due to nonlinear effects in the hearing mechanism. The secondary dips in some of the curves of Fig. 1 are due to beats between the masked tone and harmonics of the masking tone.

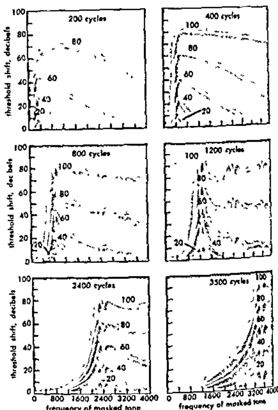


Fig. 1. Masking of pure tones by pure tones at masking frequencies of 200, 400, 800, 1200, 2400, and 3500 cps. (*From H. Fletcher, Speech and Hearing in Communication, 2d ed., Van Nostrand, 1953*)

Masking between complex sounds. Measurements of masking between complex sounds give results such as might be expected from the principles shown in Fig. 1, if subjective tones as well as the component tones are taken into account. In the masking of speech by noise, the definition of masking is often modified so that the "just-intelligible" level of speech in quiet and in noise, rather than the "just-audible" level, is used. The level for intelligibility is about 10–14 db higher than for audibility. For some special methods in the interference with speech audibility by noise, *see* PSYCHOACOUSTICS.

Masking by random noise. Critical bands are discovered in the masking of pure tones by random noise. Random noise has energy at all frequencies. It is said to be flat if the intensity level for a band of frequencies, when corrected to a 1-cycle, cen

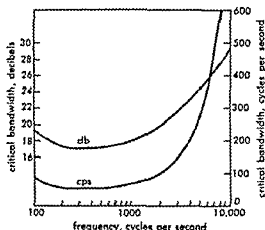


Fig. 2. Widths of critical bands, shown in cycles per second and in equivalent decibels, within the frequency range 100 cps to 10 kc. (After H. Fletcher)

tered bandwidth, is the same as that for any other band. Starting with a very narrow band of flat noise, let the band gradually be widened, keeping the intensity per cycle constant. To hear a tone whose frequency is at the center of the noise band, the intensity of the tone must be increased as the band of noise is widened. In fact, the tone will be *heard when its intensity is the same as the product* of bandwidth in cycles and noise intensity per cycle, that is, the total intensity of the noise band. However, this process continues only so far as the band is widened. When the bandwidth is increased beyond a certain critical value, the level of the center tone no longer needs to be raised to be heard. A flat noise extending over the whole hearing range has no more masking effect on a tone than a noise contained entirely within the critical band having the tone at its center. The widths of the critical bands vary with location on the frequency scale, as is shown in Fig. 2. *See* HEARING; NOISE, ACOUSTIC; SOUND. [H.K.D.]

Bibliography: H. Fletcher, *Speech and Hearing in Communication*, 2d ed., 1953; J. E. Hawkins and S. S. Stevens, The masking of pure tones and speech by white noise, *J. Acoust. Soc. Am.*, 22:6–13, 1950.

Masonry

Traditionally, constructions of stone, brick, or tile, the last being burned-clay units which are hardened by heat. Since 1900, other materials have been developed for use in masonry.

The American Standard Building Code Requirements for Masonry, approved September 28, 1953, defines masonry as "a built-up construction or combination of building units of such materials as clay, shale, concrete, glass, gypsum or stone, set in mortar or plain concrete."

The inclusion of plain concrete, that is, unreinforced concrete, is primarily for convenience in establishing building code regulations, since its properties are similar to those of unreinforced masonry. However, the commonly accepted definition of masonry, or unit masonry as it is sometimes called, is a construction of building units bonded together with mortar.

Unit masonry may be plain (unreinforced) or reinforced. Steel reinforcement is so embedded in the latter masonry that the construction will have greatly increased resistance to tensile and shearing stresses, as in walls subjected to lateral forces.

The characteristics of brick and stone masonry most responsible for their use during many centuries are compressive strength, fire resistance, and durability. Reinforced masonry has the additional property of high tensile strength.

Design of unreinforced masonry is based largely on the performance of structures over many years; also, building regulations stipulate minimum wall thickness and maximum wall heights for various types of buildings. These are included in American Standard Building Code Requirements for Masonry.

Principles of reinforced masonry design are the same as those commonly accepted for reinforced concrete, and similar formulas may be used.

Since about 1950 prefabricated masonry panels have been developed. These consist of thin slabs of reinforced masonry and are designed to be attached to steel or reinforced-concrete frames by metal anchors [H.C.P.]

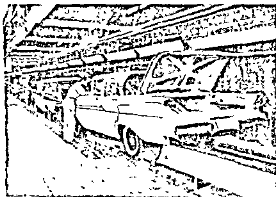
Bibliography: H. C. Plummer, *Brick and Tile Engineering*, 1950.

Mass

The quantitative or numerical measure of a body's inertia, that is, of its resistance to being accelerated.

Before this rather abstract definition is developed it is useful to consider a description of mass that, although less general, is more easily grasped intuitively. Sir Isaac Newton said that the mass of a body is the measure of the quantity of matter the body contains. This description is useful for comparing the masses of samples of a particular type of matter, say, sugar, because it correctly suggests that the mass of a sample is the basic factor in determining the extent to which the sample possesses the fundamental unchangeable properties peculiar to that type of matter. Thus a given mass of sugar, that is, a quantity of sugar that exhibits a given measure of inertia, contains a definite number of molecules and will therefore sweeten a definite number of cups of coffee or supply a definite number of calories if it is burned or eaten. Twice the mass of sugar will contain twice as many molecules.

These properties are inherent or inalienable. Their extent can be changed only by adding more sugar to the sample or taking some away, that is, only by changing mass; it will not be changed by taking the sample of sugar into a space ship or to the moon. Other quantities associated with a given sample of sugar are nonpermanent. Thus its volume can be changed by pressure and its weight changed



Automatic conveying minimizes unproductive handling costs, yet enables standardized mass-produced products to be assembled with different color combinations to give variety. (Link Belt Co.)

thousands to billions of units annually. When parts or products are produced in such large quantities, there is usually an extremely competitive market condition requiring lowest possible costs. Even those products that are marketed in relatively small quantities of a few hundred each year may use components which in themselves are manufactured in tremendously large quantities. A specific example is a standard nut or bolt purchased from a vendor who specializes in making nuts and bolts.

A mechanical part, manufactured in extremely large quantities, is made by a machine; frequently the machine is automatic. In this type of machine, accuracy of the product is largely built into the production machine. As a consequence, the machine for the making of a mass-produced part can be extremely expensive, running into the hundreds of thousands and even millions of dollars. In such situations, it is axiomatic that the requirement for highest volume with lowest possible cost warrants the expenditure of large sums for equipment.

A further characteristic of mass-produced parts and products is a high degree of standardization. A product cannot be in a state of constant change and be successfully mass produced. Changes and improvements must await major revision of the product, for example, an annual automobile model change. Great care is exercised in the initial design of such parts, particularly from the cost viewpoint. A fraction of 1 cent saved on one part is really significant if 1,000,000 parts are to be made in a year.

The effects of mass production are not all technical. In the United States, mass production, particularly of consumer goods, has contributed to a high standard of living. Other countries have large mass-production industries; however, in some nations most of the consumer goods manufactured are exported. In such instances the local standard of living is not changed appreciably. As was pointed out previously, mass production generally brings with it varying degrees of automation. The automation that lowers the unit cost of a mass product enables more people to buy the product, creates a need for new maintenance services, calls for more raw materials, and increases shipping and handling jobs both to and from the factory with consequent over-all increase in employment. See AUTOMATION; PRODUCTION ENGINEERING.

[R.S.SH.]

Mass spectroscope

An instrument used for determining the masses of atoms or molecules found in a sample of gas, liquid, or solid. It is analogous to the optical spectroscope, in which a beam of light containing various colors (white light) is sent through a prism to separate it into the spectrum of colors present. In a mass spectroscope, a beam of ions (electrically charged atoms or molecules) is sent through a combination of electric and magnetic fields so arranged that a mass spectrum is produced. If the ions fall on a photographic plate which after development shows the mass spectrum, the instrument is called a mass spectrograph; if the spectrum is allowed to sweep across a slit in front of an electrical detector which records the current, it is called a mass spectrometer.

Applications. Mass spectroscopes are used in both pure and applied science. Atomic masses can be measured very precisely. Because of the equivalence of mass and energy, knowledge of nuclear structure and binding energy of nuclei is thus gained. The relative abundances of the isotopes in naturally occurring or artificially produced elements can be determined. Thus nuclear processes occurring either in nature or in the laboratory may be investigated. Isotopic analyses of elements such as lead, argon, or strontium, which may result from the radioactive decay of other elements, are of particular interest because they make possible the determination of the geological age of the minerals from which the elements are extracted. Other geological effects which cause variations in relative abundances of isotopes may also be investigated.

The study of mass spectra of molecules provides important data on molecular structure. Mass spectrometers also make possible isotopic analyses of compounds which have reacted chemically with other compounds containing elements having artificially altered isotopic abundance ratios. Thus the instruments make possible tracer studies of chemical or biochemical reactions.

Because chemical compounds may have mass spectra as unique as fingerprints, mass spectroscopes are widely used in industries such as oil refineries, where analyses of complex hydrocarbon mixtures are required. For further information on applications, see ATOMIC WEIGHT; BETA RAYS; ISOTOPE SEPARATION (STABLE ISOTOPES).

Operation. A typical mass spectroscope has a continuously pumped vacuum chamber, commonly called the spectrometer tube, into which the gas or vapor to be investigated flows at such a rate that the equilibrium pressure in the chamber is of the order

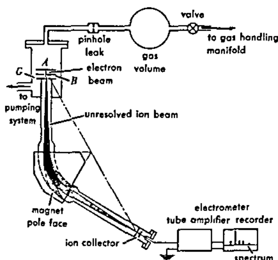


Fig. 1. Schematic drawing of mass spectrometer tube. Ion currents are in the range 10^{-10}–10^{-15} amp and require special electrometer tube amplifiers for their detection. In actual instruments, the radius of curvature of ions in a magnetic field is 10–15 cm.

of 10^{-6} mm of mercury. Figure 1 is a schematic drawing of a type of mass spectrometer tube widely used for making gas and isotope analyses. The pumping system consists of a mechanical vacuum forepump followed by either an oil or mercury diffusion pump and a cold trap maintained at dry ice or liquid air temperature. Sufficient gas of the sample to be analyzed is placed in the gas volume of several liters so that the pressure in this container is approximately 50 μ (0.05 mm Hg). A pinhole having a diameter of approximately 0.025 mm permits this gas to leak continuously into the mass spectrometer tube.

A heated tungsten filament, not shown in the diagram, produces an electron beam normal to the plane of the diagram, as shown. The electrons in the beam collide with the molecules of gas present and knock off one or more electrons, thus creating positive ions.

In a monatomic gas such as argon, multiply charged positive ions, designated as Ar^{+} (singly charged), Ar^{2+} (doubly charged), and so on, are formed. In the case of polyatomic molecules, ionized fragments may also be formed. For example, for methane, CH_4, the ions CH_4^+, CH_3^+, CH_2^+, CH^+, C^+, and H^+ are found. In some cases negative ions are also created as a result of an electron capture process (*see* ELECTRON CAPTURE). Thus for CO, in addition to the ions CO^+, C^+, O^+, the ions C^- and O^- are observed.

An electric field resulting from the application of a potential difference of several volts between A and B draws the ions through the slit in plate B. Further energy is given the ions by allowing them to fall through an electric potential of several hundreds or thousands of volts applied between plates B and G. Plate G is grounded.

The beam of ions travels downward and because of the finite width of the slits in plates B and G diverges slightly as shown. The beam passes between the poles of a magnet as indicated. The magnetic field is perpendicular to the plane of the diagram. In the magnetic field the ions experience a force at right angles to their direction of travel given by the formula

$$f = Bev \qquad (1)$$

where f is the force, B the magnetic field intensity, e the charge on the ion, and v its speed (*see* PARTICLE ACCELERATOR). This force results in a circular trajectory, the radius r of which is found by equating the force to the product of mass and acceleration according to Newton's laws of motion. Thus

$$Bev = mv^2/r \qquad (2)$$

Equation (2) may be put in a more convenient form if by equating the kinetic energy gained to the potential energy lost one expresses v in terms of the potential difference V through which the ion fell.

$$\tfrac{1}{2} mv^2 = eV \qquad (3)$$

Combining Eqs. (2) and (3), solving for r, and substituting units which are more convenient for actual calculation, one obtains

$$r = 144(mV/e)^{1/2}/B \qquad (4)$$

where r is now in cm, B in gauss, V in volts, e is the charge of the ion measured in terms of the number of electrons removed (or added) in the ionization process, and m is the mass measured in atomic mass units, that is, for hydrogen $m = 1$, for the most abundant isotope of oxygen, O^{16}, $m = 16$, and so on.

Figure 1 shows the paths of ions having three different masses. Only the intermediate group has the proper mass to reach and pass through the collector slit and be measured.

If the source of ions, the apex of the wedge-shaped field, and the collector slit all lie on a

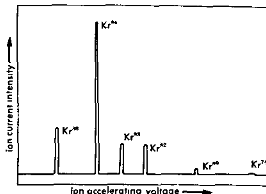
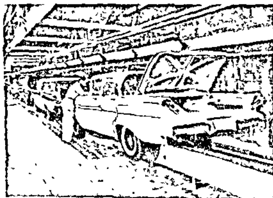


Fig. 2. Mass spectrum of krypton isotopes obtained with an instrument such as that shown in Fig. 1.



Automatic conveying minimizes unproductive handling costs, yet enables standardized mass-produced products to be assembled with different color combinations to give variety. (*Link Belt Co.*)

thousands to billions of units annually. When parts or products are produced in such large quantities, there is usually an extremely competitive market condition requiring lowest possible costs. Even those products that are marketed in relatively small quantities of a few hundred each year may use components which in themselves are manufactured in tremendously large quantities. A specific example is a standard *nut or bolt* purchased from a vendor who specializes in making nuts and bolts.

A mechanical part, manufactured in extremely large qu[illegible]
the mac[illegible]
accuracy [illegible]
duction *machine*. As a consequence, the machine for the making of a mass-produced part can be extremely expensive, running into the hundreds of thousands and even millions of dollars. In such situations, it is axiomatic that the requirement for highest volume with lowest possible cost warrants the expenditure of large sums for equipment.

A further characteristic of mass-produced parts and products is a high degree of standardization. A product cannot be in a state of constant change and be successfully mass produced. Changes and improvements must await *major revision* of the product, for example, an annual automobile model change. Great care is exercised in the initial design of such parts, particularly from the cost viewpoint. A fraction of 1 cent saved on one part is really significant if 1,000,000 parts are to be made in a year.

The effects of mass production are not all technical. In the United States, mass production, particularly of consumer goods, has contributed to a high standard of living. Other countries have large mass-production industries; however, in some nations most of the consumer goods manufactured are exported. In such instances the local standard of living is not changed appreciably. As was pointed out previously, mass production generally brings with it varying degrees of automation. The automation that lowers the unit cost of a mass product enables more people to buy the product, creates a need for new maintenance services, calls for more raw materials, and increases shipping and handling jobs both to and from the factory with consequent over-all increase in employment. See AUTOMATION; PRODUCTION ENGINEERING.

[R.S.SH.]

Mass spectroscope

An instrument used for determining the masses of atoms or molecules found in a sample of gas, liquid, or solid. It is analogous to the optical spectroscope, in which a beam of light containing various colors (white light) is sent through a prism to separate it into the spectrum of colors present. In a *mass spectroscope*, a *beam* of ions (electrically charged atoms or molecules) is sent through a combination of electric and magnetic fields so arranged that a mass spectrum is produced. If the ions fall on a photographic plate which after development shows the mass spectrum, the instrument is called a mass spectrograph; if the spectrum is allowed to sweep across a slit in front of an electrical detector which records the current, it is called a mass spectrometer.

Applications. Mass spectroscopes are used in both pure and applied science. Atomic masses can be measured very precisely. Because of the equivalence of mass and energy, knowledge of nuclear structure and binding energy of nuclei is thus gained. The relative abundances of the isotopes in naturally occurring or artificially produced elements can be determined. Thus nuclear processes occurring either in nature or in the laboratory may be investigated. Isotopic analyses of elements such as lead, argon, or strontium, which may result from the radioactive decay of other elements, are of particular interest because they make possible the determination of the geological age of the minerals from which the elements are extracted. Other geological effects which cause variations in relative abundances of isotopes may also be investigated.

The study of mass spectra of molecules provides important data on molecular structure. Mass spectrometers also make possible isotopic analyses of compounds which have reacted chemically with other compounds containing elements having artificially altered isotopic abundance ratios. Thus the instruments make possible tracer studies of chemical or biochemical reactions.

Because chemical compounds may have mass spectra as unique as fingerprints, mass spectroscopes are widely used in industries such as oil refineries, where analyses of complex hydrocarbon mixtures are required. For further information on applications, *see* ATOMIC WEIGHT; BETA RAYS; ISOTOPE SEPARATION (STABLE ISOTOPES).

Operation. A typical mass spectroscope has a continuously pumped vacuum chamber, commonly called the spectrometer tube, into which the gas or vapor to be investigated flows at such a rate that the equilibrium pressure in the chamber is of the order

Handling Society has classified equipment as (1) conveyor, (2) cranes, elevators, hoists, (3) positioning, weighing, and control equipment, (4) industrial vehicles, (5) motor vehicles, (6) railroad cars, (7) marine carriers, (8) aircraft, and (9) containers and supports.

Every materials handling problem starts with the material—its dimensions, its nature, its characteristics. Engineers who fail to start here usually end up trying to justify equipment, rather than achieving safe and economical movement of the material. The quantity to be moved—both in total and in rate of moving desired—next most affects the type of handling. Then comes the sequence of operations or the routing. Basically, this what, when (how much and how often), and where is the minimum information needed to evaluate or determine any handling system or equipment.

Materials handling is both a planning and an operating activity. These two activities are generally separated in industry; an analytical group designs or selects the system or equipment and the operating group puts it to use. See PLANT FACILITIES. [R.M.]

Materials handling machines

Devices used for handling materials in an industrial or a distribution activity. The equipment moves ...

do not include the means employed to control the flow of fluids.

Many different types of machines result from combinations and permutations of the following factors: (1) the route over which the product is moved may be fixed or variable; (2) the path of travel may be horizontal, inclined, declined, or vertical; (3) motion may be imparted to the product manually, by the force of gravity, by compressed air, by suction, by vibration, or by power-actuated components of the machine; (4) the motion may be continuous or intermittent (reciprocating); (5) the product may be supported or carried suspended during the handling operation.

Based on their most common characteristics, materials handling machines can be grouped in six broad categories (see BULK-HANDLING MACHINES; CONVEYING MACHINES; ELEVATING MACHINES; HOISTING MACHINES; INDUSTRIAL TRUCKS; MONORAIL). Improvements in handling techniques stem from the wide adoption by industry of palleting and of the forklift truck during World War II. These innovations have produced far-reaching effects. Among these are radical changes in plant layout, elevator design for multistory operations, and the increasing trend to single-story facilities.

Automation in the sense of feedback control and advanced mechanization in the fabrication and transfer of products from one operation to the next brings together two major industrial technologies (see PRODUCTION ENGINEERING; TRANSPORTATION ENGINEERING). Electronic data processing devices facilitate the compiling of inventory records and the handling of orders. Use of photoelectric tubes for counting and controlling the action of doors, conveyors, and other materials handling machines is another example of how electronic techniques are being applied. One- and two-way radio and, more recently, television improve the communication in plants and yards. [D O.H.]

Bibliography: H. A. Bolz and G. E. Hagemann (eds.), *Materials Handling Handbook*, 1958; D. O. Haynes, *Materials Handling Equipment*, vol. 1, 1957; D. O. Haynes, *Materials Handling Applications*, vol. 2, 1958.

Maternal influence

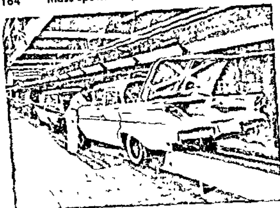
The determination of characters of the progeny by the maternal parent. This effect may be mediated by predetermination of the ovum or, in mammals, by effects of the uterine environment or of the milk. Both the genetic constitution of the mother and environmental conditions affecting the mother may be responsible.

In the moth *Ephestia*, the gene *a* blocks the oxidation of tryptophan to kynurenin and consequently the formation of pigments derived from kynurenin. In a^+a^+ and a^+a females 3-hydroxykynurenin is stored in the ovum. The *aa* larvae from a cross ♀ a^+a × ♂ aa can use this substance for pigment formation and have pigmented ommatidia and skins. Similarly, in the snail *Limnaea* the direction of coiling of the offspring's shell is determined by the genes of the mother, probably by determination of the position of the spindle during the first cleavage division.

Body size and the number of lumbar vertebrae in the mouse are affected by the uterine environment. Nutritional deficiencies in the mouse may lead to well-defined malformations, depending on the type of deficiency and the strain used. In the rat, progeny of mothers subjected to anxiety-producing situations show behavioral differences compared to control animals.

Maternal effects may be mediated through transmission of nonreproducing substances as in *Ephestia* or by transmission of self-reproducing organisms, for example, symbiotic bacteria transmitted through the ovum of insects, or rickettsia deposited in the ovum of ticks. Similar are the milk-factor particles in mouse strain C3H which are transmitted through the milk of the mother and induce the formation of mammary cancers in the progeny. The milk-factor particles are of submicroscopic size, multiply in the cells, and depend for multiplication on genes of the host. Except for their mode of transmission, they resemble the cytoplasmic particles responsible for the killer effect in *Paramecium*. See CYTOPLASMIC INHERITANCE; RICKETTSIOSES; TICK.

A more complex type of interaction between mother and developing offspring is exemplified by



Automatic conveying minimizes unproductive handling costs, yet enables standardized mass-produced products to be assembled with different color combinations to give variety. (Link Belt Co)

thousands to billions of units annually. When parts or [illegible] such large quantities, the [illegible] market co[illegible]s. Even those products that are marketed in relatively small quantities of a few hundred each year may use components which in themselves are manufactured in tremendously large quantities A specific example is a standard nut or bolt purchased from a vendor who specializes in making nuts and bolts.

A mechanical part, manufactured in extremely large quantities, is made by a machine; frequently the machine is automatic. In this type of machine, accuracy of the product is largely built into the production machine. As a consequence, the machine for the making of a mass-produced part can be extremely expensive, running into the hundreds of thousands and even millions of dollars. In such situations, it is axiomatic that the requirement for highest volume with lowest possible cost warrants the expenditure of large sums for equipment.

A further characteristic of mass-produced parts and products is a high degree of standardization. A product cannot be in a state of constant change and be successfully mass produced. Changes and improvements must await *major revision* of the product, for example, an annual automobile model change Great care is exercised in the initial design of such parts, particularly from the cost viewpoint. A fraction of 1 cent saved on one part is really significant if 1,000,000 parts are to be made in a year.

The effects of mass production are not all technical. In the United States, mass production, particularly of consumer goods, has contributed to a high standard of living. Other countries have large mass-production industries; however, in some nations most of the consumer goods manufactured are exported. In such instances the local standard of living is not changed appreciably. As was pointed out previously, mass production generally brings with it varying degrees of automation. The automation that lowers the unit cost of a mass product enables more people to buy the product, creates a need for new maintenance services, calls for more raw materials, and increases shipping and handling jobs both to and from the factory with consequent over-all increase in employment. See AUTOMATION; PRODUCTION ENGINEERING.

[R.S.H.]

Mass spectroscope

An instrument used for determining the masses of atoms or molecules found in a sample of gas, liquid, or solid. It is analogous to the optical spectroscope, in which a beam of light containing various colors (white light) is sent through a prism to separate it into the spectrum of colors present. In a mass spectroscope, a beam of ions (electrically charged atoms or molecules) is sent through a combination of electric and magnetic fields so arranged that a mass spectrum is produced. If the ions fall on a photographic plate which after development shows the mass spectrum, the instrument is called a mass spectrograph; if the spectrum is allowed to sweep across a slit in front of an electrical detector which records the current, it is called a mass spectrometer.

Applications. Mass spectroscopes are used in both pure and applied science. Atomic masses can be measured very precisely. Because of the equivalence of mass and energy, knowledge of nuclear structure and binding energy of nuclei is thus gained. The relative abundances of the isotopes in naturally occurring or artificially produced elements can be determined. Thus nuclear processes occurring either in nature or in the laboratory may be investigated. Isotopic analyses of elements such as lead, argon, or strontium, which may result from the radioactive decay of other elements, are of particular interest because they make possible the determination of the geological age of the minerals from which the elements are extracted. Other geological effects which cause variations in relative abundances of isotopes may also be investigated.

The study of mass spectra of molecules provides important data on molecular structure. Mass spectrometers also make possible isotopic analyses of compounds which have reacted chemically with other compounds containing elements having artificially altered isotopic abundance ratios. Thus the instruments make possible tracer studies of chemical or biochemical reactions.

Because chemical compounds may have mass spectra as unique as fingerprints, mass spectroscopes are widely used in industries such as oil refineries, where analyses of complex hydrocarbon mixtures are required. For further information on applications, see ATOMIC WEIGHT; BETA RAYS; ISOTOPE SEPARATION (STABLE ISOTOPES).

Operation. A typical mass spectroscope has a continuously pumped vacuum chamber, commonly called the spectrometer tube, into which the gas or vapor to be investigated flows at such a rate that the equilibrium pressure in the chamber is of the order

of permanently frozen ground. *See* SOLIFLUCTION.

Rapid flowage, characteristic only of certain regions and typically in individual occurrences rather than of wide areal extent, includes earthflow, mudflow, and debris-avalanche.

Earthflow is particularly common in areas of weathered shale bedrock or clay. In hillside earthflows (Fig. 3), an area of soil and mantle rock to a depth of several feet is unable when saturated to maintain its form on the slope and flows a short distance downhill, often confined beneath a tight sod. Movement slows and stops when the bulging toe cracks and some of the water escapes. Because earthflow removes support from the area directly upslope, a landslide of slump type usually results. Earthflow may be moderately rapid or slow and intermittent over a long period.



Fig. 2. View showing hillside creep in a tributary of the Yukon, Yukon Territory. (USGS)



Fig. 3. Freshly formed earthflow near Berkeley, California. (USGS)



Fig. 4. Mudflow as viewed from its source in far background to Lake San Cristobal, Hinsdale County, Colorado. (USGS)

Mudflow is a phenomenon of semiarid lands, subalpine mountains, or volcanic slopes. Mudflows usually follow a drainageway and often recur along the course of previous mudflows (Fig. 4). The material is wetter and the movement faster than earthflow. Large blocks of rock carried in a thick mudflow can be moved far out from a mountain front.

Debris-avalanche is the equivalent of the mudflow in mountains of humid regions. It usually involves the entire thickness of the soil mantle down to bedrock but does not follow a previous avalanche track. Deposits are heaps, fans, or valley trains of heterogenous rock debris, usually cluttered with trees and other vegetation.

Landslide includes sliding and falling movements of relatively dry masses of earth and rocks. The term is sometimes used broadly to cover all forms of rapid mass movement. *See* LANDSLIDE.

Subsidence is a downward movement without any forward or outer component, as in the settling of ground over mines. [C.F.S.S.]

Bibliography: C. F. S. Sharpe, *Landslides and Related Phenomena*, 1938.

Massif

A block of the earth's crust commonly consisting of crystalline gneisses and schists, the textural appearance of which is generally markedly different from that of the surrounding rocks. Common usage indicates that a massif has limited areal extent and considerable topographic relief. Structurally, a massif may form the core of an anticline or may be a block bounded by faults or even unconformities. In any case, during the final stages of its development a massif acts as a relatively homogeneous tectonic unit, which to some extent controls the structures that surround it. Numerous complex internal structures may be present; many of these are not related to its development as a massif but are the mark of previous deformations. *See* TECTONIC PATTERNS. [P.H.O.]

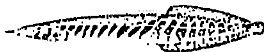
Mass-transfer operation

An operation in chemical engineering that involves transfer of material from one phase to another, or from one place to another within a single phase. Mass transfer may occur for various purposes, for example, to effect a chemical reaction, to obtain a separation of components, or to obtain uniform distribution of material within a phase. The most common types are (1) fluid-fluid transfer, occurring typically in absorption, distillation, liquid-liquid extraction, and (2) fluid-solid

fer, occurring typically in crystallization and gas adsorption. The subject of mass transfer deals generally with the factors which determine rate of interphase material transfer and with the influence of transfer rate on equipment design and performance. Processes involving mass transfer are often designated as diffusional operations. *See* ADSORPTION; COUNTERCURRENT MASS-TRANSFER OPERATION; CRYSTALLIZATION; DIALYSIS; DIFFUSION IN GASES AND LIQUIDS; DISTILLATION; DRYING; GAS ABSORPTION OPERATIONS; HUMIDIFICATION; ION EXCHANGE; LEACHING; SOLVENT EXTRACTION; SUBLIMATION. [C.R.W.]

Mastacembeliformes

A small order of actinopterygian fishes derived from perciform forebears, the anguilliform body having been developed independently from that of the true eels. This group is also known as the Opisthomi. Its characteristics include the upper jaw bordered by the premaxillae, median fins nearly continuous, spines usually present in front of the dorsal and anal fins, no air duct, no posttemporal bone, the pectoral girdle suspended by a ligament from the vertebral column, a pectoral fin but no pelvic fin, snout usually produced and ending in a fleshy tentacle, 70–95 vertebrae, and scales either present or absent.



Mastacembelid eel, Mastacembelus circumcinctus. (After H. M. Smith, The Fresh-Water Fishes of Siam or Thailand, U.S. Natl. Museum Bull. 188, 1945)

There are 2 families, one containing a single small species from Burma, the other with 2 genera and about 50 species that inhabit fresh and brackish waters of Africa and southern Asia. No fossil forms are known. *See* ACTINOPTERYGII. [R.M.B.]

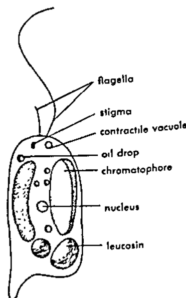
Mastigophora

A subphylum of the protozoa also known as the Flagellata. The flagellates comprise a large and heterogeneous group of Protozoa, varying in size from Noctiluca, 1500 μ in diameter, to monads, barely 3 μ long. The common morphological flagellate type is spherical to cylindrical on an anteroposterior axis. They are probably more nearly representative of the primitive protozoan type than Sarcodina or Ciliophora, but exhibit a more evident relationship to the former. Despite their common morphological plan, based on the flagellum as a means of locomotion, Flagellata are very diverse in shape, colony formation, internal structure, external shell or test, color, physiology, reproduction, and choice of environment.

The group exhibits marked plantlike characteristics, so that texts and references in botany always treat at least some of them. Some workers include all colorless flagellates in the algae; on the other hand, flagellates display many distinctly animal features and hence sometimes are treated as Protozoa. Generally, flagellates are regarded as a sort of connecting link between the plant and animal kingdoms.

Morphology. Mastigophora possess one common feature, locomotion by means of one or more whiplike protoplasmic extrusions termed flagella. A few have secondarily lost this organelle. In some it is single and directed anteriorly; in a few it is directed posteriorly. Others have one anterior and one trailing flagellum, but there are numerous combinations such as one anterior and two trailing (*Amphimonas*), or up to eight anterior (*Polyblepharides*). Dinoflagellates customarily have two laterally inserted flagella, one posteriorly directed, the other encircling the cell. *Trepomonas* and its relatives have two lateral sets, and the Hypermastigina, symbiotic in termites and wood roaches, have large numbers of very long flagella, mostly lateral. The flagella may be equal or subequal. Usually the flagellar equipment is handed on to one daughter cell at division, a new set being produced in the other. In some species (*Mastigamoeba reptans*) the flagellum may be temporarily withdrawn, at which time the organism moves by ameboid action. Flagella are usually few in number, except in Hypermastigina.

Body form. The flagellate body is typically monaxial and elongate to some degree. Cells are practically spherical in *Monas*, but ovoid, cordiform, pyriform, fusiform, acicular, tubular, or flattened cells are more common. Flattened species typically glide along a surface rather than swim. A particular shape is normally maintained by *Ochromonas* (see illustration), whereas *Mastigamoeba* and some others form pseudopodia, and certain species



Ochromonas ludibunda.

(*Euglena agilis*) undergo frequent distortions termed euglenoid or metabolic changes of shape.

Structurally the flagellate cell is not simple. The cytoplasm is sometimes quite vacuolated (*Collodictyon* or *Trepomonas*) and occasionally colored, although color is generally confined to chromatophores when these are present. One or more contractile vacuoles may be located near the flagellar base.

Colony formation. Most flagellates are single-celled but colony formation is frequent. Cells may be naked (*Oicomonas*), enclosed in a thick cellulose cell wall or pellicle (*Euglena spirogyra*), or in a chitinous, calcareous, or silicious test, the lorica or shell (*Trachelomonas*, *Distephanus speculum*). The lorica may be smooth, spiny, sculptured, or with attached foreign bodies (*Urceolus sabulosus*). Tests may be clear or colored by impregnated iron. Colony forms are variously flattened (*Platydorina*), spherical (*Volvox*), irregular (*Cladospongia*), linear (*Desmarella*), catenate (*Gonyaulax*), tubular (*Stelexomonas*), or dendroid. Colony walls are usually gelatinous in texture but vary Single cells are frequently attached by a stalk to some substrate, especially those living in cup-shaped tests. The cell proper rarely shows protuberant parts other than flagella and pseudopodia, but the Protomonadina include many species showing one (*Monosiga*, *Proterospongia*), or two (*Dicraspedella*) protoplasmic collars ringing the base of the flagellum.

Flagellum. The flagellum contains a structurally complex longitudinal axial fibril or axoneme Along its length numerous short lateral hairlike mastigonemes occur which can be seen only in special preparations. The flagellum originates in a kinetic body, the blepharoplast, usually close to the outer surface of the cell. Blepharoplasts for each flagellum are connected by fibrils termed paradesmoses. Another fibril, the rhizoplast, may extend from the blepharoplast to the centriole or to the nucleus. There may be still other structures in the flagellar complex, demonstrable by special techniques. The whole flagellar apparatus produces coordinated movement. Red eye spots (stigmas) may occur near the base of the flagellum. They are considered photoreceptors, and are sometimes accompanied by a lenslike structure; the whole is important in phototactic orientation. *See* PHOTORECEPTION.

Nucleus. Nuclei are single except in the distomates (*Hexamitus*, *Giardia*) and each undergoes mitosis in division. Some contain endosomes of unknown function but do not contribute directly to chromosome formation. Others contain nucleoli which apparently are identical with, or contribute to, chromatin. Dinoflagellate nuclei are large and vesicular; each chromosome apparently retains its identity during the interphase. In the Euglenoidina the chromatin granules are typically scattered with one or more centrally located endosomes. Nuclei are large in Chloromonadida and are small in Chrysomonadida, Cryptomonadida, Phytomonadida, Zoomastigophorea, Rhizomastigida, and some Protomonadida. In these the chromatin tends to peripheral location with central endosomes.

Nuclear and cell division. In species whose nuclear division is mitotic, details vary, but definite and species-characteristic chromosome numbers occur. In some cases (*Bodopsis godboldi*, *Dimorpha mutans*), well-defined spindles, within or without the original nuclear membrane, are present. Blepharoplasts function as centrioles. Other mitoses may lack spindles or centrioles (*Oxyrrhis marina*, *Entosiphon sulcatum*). Mitosis in Hypermastigida is extremely complicated. Most flagellates are haploid; gamete formation is the exception although sexual processes are known for Phytomonadina. *Pleodorina californica* has visibly different somatic and reproductive cells.

Cell division is generally longitudinal binary fission which sometimes occurs while the organism is encysted, but more often while it is active. Occasionally, multiple fission occurs as in *Chlorogonium* whose trophozoites are sometimes seen to be divided into eight daughter cells. *See* MITOSIS.

Cell inclusions. Flagellates contain various cell inclusions, such as mitochondria and Golgi apparatus. Conspicuous inclusions are chromatophores and reserve food materials. Chromatophores are discoid, cup-shaped, or irregular networks containing one or more of the chlorophylls and often carotene, xanthophyll, or other pigments. Green is characteristic for Phytomonadida and Euglenida, yellow for Chrysomonadida. Cryptomonadida and Dinoflagellida are blue, brown, olive green, pink, or red. Chromatophores may contain pyrenoids as centers of starch formation. Other supposed reserve food inclusions are paramylum (Euglenida); leucosin and volutin (Chrysomonadida); and glycogen (Chloromonadida). Lipids are widely distributed. Occasionally hematochrome granules are sufficiently abundant to color the cell, as when blooms of *Euglena haematoides* redden the surfaces of ponds in hot summer weather. *See* CAROTENOID; CHLOROPHYLL; CHROMATOPHORE; GOLGI COMPONENT; MITOCHONDRIA.

Nematocysts, trichocysts, and trichites occur in flagellates. In *Polykrikos* (Dinoflagellida) nematocysts lie free in the cytoplasm, but their function is unknown. Trichocysts in *Polykrikos*, in *Euglena* species, and in *Chilomonas* may be discharged as long threads by various stimuli, but here too their function is open to speculation.

Nutrition. Nutrition is holophytic when food is manufactured from inorganic substances by those species containing chlorophyll; saprozoic, when dissolved organic materials are absorbed and resynthesized; and holozoic, when solid food is ingested. *Euglena gracilis*, which is holophytic in the light, may become saprozoic in the dark. *Euglena acus* may become saprozoic in the light in a rich organic medium. *Ochromonas ludibunda* apparently is simultaneously holophytic and holozoic. Fl

lata perhaps more than any other group vary widely in nutritive processes from those involving intake of specific chemical entities to those of synthesis. Their requirements of trace elements, vitamins, and amino acids vary almost from species to species, as shown by precise experimental investi- [illegible] flag- [illegible] prob- [illegible] richo- *phorum*), but some (*Gonyostomum semen*) appear to be obligate aerobes, others (*Trepomonas rotans*) obligate anaerobes Carbon dioxide is excreted through the contractile vacuole as is oxygen. Dense blooms of *Chlamydomonas* or *Euglena* will supersaturate surface waters with oxygen within a few hours after daybreak. Excretion of other metabolites such as antibiotics is demonstrable. Circulation by cyclosis is sometimes observable.

Ecology. Flagellates adjust to wide ranges of pH, osmotic pressure, light, and temperature but optima are found upon investigation. *Euglena mutabilis* grows at pH 0.9 and 7.2, but densest blooms occur about pH 3 5–4.0. Some few flagellate species occur in both fresh and sea water, but cannot be transferred directly from fresh to sea water.

Almost any aqueous ecological niche contains Flagellata Dense fresh-water blooms of *Euglena* or *Platymonas* are matched by blooms of *Eutreptia* or *Gymnodinium breve* in the ocean. *Trentonia flagellata* is the only green flagellate occurring abundantly in Warm Mineral Springs, an anaerobic sulfur spring of southern Florida. Euglenida and Phytomonadida grow profusely in barnyard pools, Chloromonadida in weakly acid cedar and cypress swamp water, and Phytomastigida and Cryptomonadida in lower nutrient levels of oceans, streams, and lakes. Dinoflagellida abound in the oceans; Chrysomonadida tend to live in clean water; Zoomastigida and Rhizomastigida favor organically polluted water. There is much overlapping because many are ubiquitous.

Flagellata are near the base of the food pyramid, probably next to bacteria. Their varied synthetic abilities, fast reproductive rate, and huge numbers in both fresh and salt water compensate for their usual small size. The oceanic blooms which kill fish and other animals are principally flagellates, but green flagellates reaerate polluted water. They often cause tastes and odors in potable water, but they readily attack organic matter in natural water as do bacteria. They include many parasites dangerous to man and animals and are themselves frequently parasitized. *See* PROTOZOA; *see also* ALGAE; FOOD CHAIN. [J.B.L.]

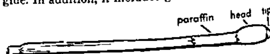
Mastitis (cows)

Inflammation of breast tissues. In infants, mastitis from temporary hormonal inbalance affects both sexes and disappears spontaneously. Chronic mastitis of women involves anatomic changes which lead to secondary infection. The microorganism *Streptococcus hemolyticus* causes mastitis in cows. Serologic Group B (bovine) strains can infect a herd without human sickness. Group A (human) streptococci are introduced accidentally to the cow's udder by a handler. This infected milk has caused human epidemics of septic sore throat. Mastitis of cows concerns veterinarians, dairymen, microbiologists, and physicians. Milk favors bacterial growth, and the differentiation of streptococci by host, the immediate source, is not absolute. *See* LANCEFIELD DIFFERENTIATION SCHEME; MILK; STREPTOCOCCUS. [P.L.B.]

Match

A short piece of wood or other material at the end of which has been placed a substance that can be ignited by friction.

Matches are now of two types, strike-anywhere and safety. The strike-anywhere match contains potassium chlorate ($KClO_3$) and phosphorus sesquisulfide (P_4S_3) as the main ingredients, with added quantities of iron oxide, zinc oxide, and glue. In addition, it includes ground glass to create



Strike-anywhere match.

friction. Safety matches involve two separate compositions, the first coated on the matchstick and the second on the side of the container. The first contains potassium chlorate, potassium dichromate, manganese dioxide, sulfur, iron oxide, ground glass, and glue. The second contains red phosphorus, antimony sulfide, and an abrasive. The heat produced by the ignition of a minute amount of the red phosphorus when the match is struck is sufficient to ignite the matchhead. The wood or cardboard of the matchstick is coated with paraffin to give a longer and more ready flame and is impregnated with borax to prevent afterglow.

The first matches contained white phosphorus, which ignites at a very low temperature (50°C in air). Red phosphorus, an allotropic form of the element discovered much later, ignites at a considerably higher temperature (about 240°C in air). Extensive use of white phosphorus in matches after they were invented in the nineteenth century led to the discovery of a pernicious bone degeneration in factory workers. Phosphorus vapor apparently enters through dental caries and leads to a deterioration of the jaw bone called phossy jaw. Proscriptions against its use have now eliminated white phosphorus from matches, and its place has been taken by the phosphorus sesquisulfide. Red phosphorus is not poisonous. *See* COMBUSTION; FLAME; PHOSPHORUS; PYROTECHNICS. [W.E.G.]

Materials handling

Loading, moving, and unloading of ore from mine to factory and of garments within a loft are examples of materials handling. There are hundreds

of different ways of handling materials. These are generally classified according to the type of equipment used. For example, the American Materials Handling Society has classified equipment as (1) conveyor, (2) cranes, elevators, hoists, (3) positioning, weighing, and control equipment, (4) industrial vehicles, (5) motor vehicles, (6) railroad cars, (7) marine carriers, (8) aircraft, and (9) containers and supports.

Every materials handling problem starts with the material—its dimensions, its nature, its characteristics. Engineers who fail to start here usually end up trying to justify equipment, rather than achieving safe and economical movement of the material. The quantity to be moved—both in total and in rate of moving desired—next most affects the type of handling. Then comes the sequence of operations or the routing. Basically, this *what*, when (how much and how often), and where is the minimum information needed to evaluate or determine any handling system or equipment.

Materials handling is both a planning and an *operating activity. These two activities are generally* separated in industry; an analytical group designs or selects the system or equipment and the operating group puts it to use. *See* PLANT FACILITIES. [R.M.]

Materials handling machines

Devices used for handling materials in an industrial or a distribution activity. The equipment moves products as discrete articles, in suitable containers, or as solid bulk materials which are relatively *free-flowing. Materials handling machines* do not include the means employed to control the flow of fluids.

Many different types of machines result from combinations and permutations of the following factors: (1) the route over which the product is moved may be fixed or variable; (2) the path of travel may be horizontal, inclined, declined, or vertical; (3) motion may be imparted to the product manually, by the force of gravity, by compressed air, by suction, by vibration, or by power-actuated components of the machine; (4) the motion may be continuous or intermittent (reciprocating); (5) the product may be supported or carried suspended during the handling operation.

Based on their most common characteristics, materials handling machines can be grouped in six broad categories (*see* BULK-HANDLING MACHINES; CONVEYING MACHINES; ELEVATING MACHINES; HOISTING MACHINES; INDUSTRIAL TRUCKS; MONORAIL). Improvements in handling techniques stem from the wide adoption by industry of palleting and of the forklift truck during World War II. These innovations have produced far-reaching effects. Among these are radical changes in plant layout, elevator design for multistory operations, and the increasing *trend to single-story* facilities.

Automation in the sense of feedback control and advanced mechanization in the fabrication and transfer of products from one operation to the next brings together two major industrial technologies (*see* PRODUCTION ENGINEERING; TRANSPORTATION ENGINEERING). Electronic data processing devices facilitate the compiling of inventory records and the handling of orders. Use of photoelectric tubes for counting and controlling the action of doors, conveyors, and other materials handling machines is another example of how electronic techniques are being applied. One- and two-way radio and, more recently, television improve the communication in plants and yards. [D.O.H.]

Bibliography: H. A. Bolz and G. E. Hagemann (eds.), *Materials Handling Handbook*, 1958; D. O. Haynes, *Materials Handling Equipment*, vol. 1, 1957; D. O. Haynes, *Materials Handling Applications*, vol. 2, 1958.

Maternal influence

The determination of characters of the progeny by the maternal parent. This effect may be mediated by predetermination of the ovum or, in mammals, by effects of the uterine environment or of the milk. *Both the genetic constitution of the mother and en*vironmental conditions affecting the mother may be responsible.

In the moth *Ephestia*, the gene *a* blocks the oxidation of tryptophan to kynurenin and consequently the formation of pigments derived from kynurenin. In a^+a^+ and a^+a females 3-hydroxykynurenin is stored in the ovum. The *aa* larvae from a cross ♀ a^+a × ♂ *aa* can use this substance for pigment formation and have pigmented ommatidia and skins. Similarly, in the snail *Limnaea* the di*rection of coiling of the offspring's shell* is determined by the genes of the mother, probably by determination of the position of the spindle during the first cleavage division.

Body size and the number of lumbar vertebrae in the mouse are affected by the uterine environment. Nutritional deficiencies in the mouse may lead to well-defined malformations, depending on the type of deficiency and the strain used. In the rat, progeny of mothers subjected to anxiety-producing situations show behavioral differences compared to control animals.

Maternal effects may be mediated through transmission of nonreproducing substances as in *Ephestia* or by transmission of self-reproducing organisms, for example, symbiotic bacteria transmitted through the ovum of insects, or rickettsia deposited in the ovum of ticks. Similar are the milk-factor particles in mouse strain C3H which are transmitted through the milk of the mother and induce the formation of mammary cancers in the progeny. The milk-factor particles are of submicroscopic size, multiply in the cells, and depend for multiplication on genes of the host. Except for their mode of transmission, they resemble the cytoplasmic particles responsible for the killer effect in *Paramecium*. *See* CYTOPLASMIC INHERITANCE; RICKETTSIOSES; TICK.

A more complex type of interaction between mother and developing offspring is exemplified by

the Rh antigens in man. These are erythrocyte or red blood cell antigens determined by a series of multiple alleles or possibly pseudoalleles. If a woman belonging to one of certain antigenic types (Rh negative) carries an embryo which due to [illegible] duce antibodies against the embryo. If the antibody titer is high, which may occur particularly after repeated immunizations, the antibodies may interfere with the differentiation of the embryo, producing the frequently fatal hemolytic disease of the newborn, erythroblastosis fetalis. *See* Blood groups; Genetics. [E.W.C.]

Mathematical notation, contemporary

Most special symbols and notational conventions are seen only within some specialized branch of mathematics; also, it has become more and more usual for each book and paper to define its own set of notations adapted to its own purpose. The following list emphasizes those notations which are used in several branches of mathematics and which authors may expect readers to understand without explanation. The left column sometimes contains the symbol (or several versions of the symbol), sometimes an instance of its use. The interpretation in the right column generally does not include a definition. Some notations are here, not because they have fixed, specific meanings, but only to illustrate current usage

Mathematical logic

p, q, $P(x)$	Sentences, propositional functions, propositions
$\neg p$, $\sim p$, non p, Np, $/$	Negation, read "not p" ($\neq$: read "not equal")
$p \vee q$, $p + q$, Apq	Disjunction, read "p or q," "p, q," or both
$p \wedge q$, $p \quad q$, $p \& q$, Kpq	Conjunction, read "p and q"
$p \rightarrow q$, $p \supset q$, $p \Rightarrow q$, Cpq	Implication, read "p implies q" or "if p then q"
$p \leftrightarrow q$, $p \equiv q$, $p \Leftrightarrow q$, Epq, p iff q	Equivalence, read "p is equivalent to q" or "p if and only if q"
n.a.s c	Read "necessary and sufficient condition"
( ), [ ], { }, ...	Parentheses
$\forall$, $\forall$, Σ	Universal quantifier, read "for all" or "for every"
$\exists$, $\exists$, Π	Existential quantifier, read "there is a" or "there exists"
$\vdash$	Assertion sign ($p \vdash q$: read "q follows from p"; $\vdash p$: read "p is or follows from an axiom," or "p is a tautology"
0, 1	Truth, falsity (values)
$=$	Identity
$\overset{\text{Df}}{=}$, $\overset{\text{df}}{=}$, $\underset{\text{df}}{=}$, $\equiv$	Definitional identity
$\blacksquare$	"End of proof"; "QED"

See Logic

Set theory, relations, functions

X, Y	Sets
$x \in X$	x is a member of the set X
$x \notin X$	x is not a member of X
$A \subset X$, $A \subseteq X$	Set A is contained in set X
$A \not\subset X$, $A \not\subseteq X$	A is not contained in X
$X \cup Y$, $X + Y$	Union of sets X and Y
$X \cap Y$, $X \cdot Y$	Intersection of sets X and Y
$+$, $\dotplus$, $\bigcirc$	Symmetric difference of sets
$\bigcup X_i$, ΣX_i	Union of all the sets X_i
$\bigcap X_i$, ΠX_i	Intersection of all the sets X_i
ϕ, 0, Λ	Null set, empty set
X', $\complement X$, CX	Complement of the set X
$X - Y$, $X \backslash Y$	Difference of sets X and Y
$\hat{x}(P(x))$, $\{x \mid P(x)\}$, $\{x : P(x)\}$	The set of all x with the property P
(x,y,z), $\langle x,y,z \rangle$	Ordered set of elements x, y, and z; to be distinguished from (x,z,y), for example
$\{x,y,z\}$	Unordered set, the set whose elements are x, y, z, and no others
$\{a_1, a_2, \ldots, a_n\}$, $\{a_i\}_{i=1,2,\ldots,n}$, $\{a_i\}_{i=1}^{n}$	The set whose members are a_i, where i is any whole number from 1 to n
$\{a_1, a_2, \ldots\}$, $\{a_i\}_{i=1,2,\ldots}$, $\{a_i\}_{i=1}^{\infty}$	The set whose members are a_i, where i is any positive whole number
$X \times Y$	Cartesian product, set of all (x,y) such that $x \in X$, $y \in Y$
$\{a_i\}_{i \in I}$	The set whose elements are a_i, where $i \in I$
xRy, $R\{x,y\}$	Relation
$\equiv$, $\cong$, $\sim$, $\simeq$	Equivalence relations, for example, congruence
$\geqq$, $\geq$, $>$, $\succ$, $\gg$, $\leqq$, $\leq$, $<$	Transitive relations, for example, numerical order
$f: X \rightarrow Y$, $X \xrightarrow{f} Y$, $X \rightarrow Y$, $f \in Y^X$	Function, mapping, transformation
f^{-1}, $\overset{-1}{f}$, $X \xleftarrow{f^{-1}} Y$	Inverse mapping
$g \circ f$	Composite functions: $(g \circ f)(x) = g(f(x))$
$f(X)$	Image of X by f
$f^{-1}(X)$	Inverse-image set, counter image
1-1, one-one	Read "one-to-one correspondence"
$X \xrightarrow{f} Y$ $\phi\downarrow \quad \downarrow\psi$ $W \xrightarrow{g} Z$	Diagram: the diagram is commutative in case $\psi \circ f = g \circ \phi$
$f \mid A$	Partial mapping, restriction of function f to set A
$\bar{\bar{X}}$, card X, $\lvert X \rvert$	Cardinal of the set A
$\aleph_0$, d	Denumerable infinity
$\mathfrak{c}$, c, $2^{\aleph_0}$	Power of continuum
ω	Order type of the set of positive integers
σ-	Read "countably"

See Conformal mapping; Graph theory; Group theory; Set theory; Topology.

Number, numerical functions

1.4; 1,4; 1·4	Read "one and four-tenths"
1(1)20(10)100	Read "from 1 to 20 in intervals of 1, and from 20 to 100 in intervals of 10"
const	Constant
$A \geqq 0$	The number A is nonnegative, or, the matrix A is positive definite, or, the matrix A has nonnegative entries
$x\|y$	Read "x divides y"
$x \equiv y \bmod p$	Read "x congruent to y modulo p"
$a_0 + \frac{1}{a_1 +} \frac{1}{a_2 +} \cdots$, $a_0 + \frac{1\|}{\|a_1} + \cdots$	Continued fractions
$[a,b]$	Closed interval
$[a,b)$, $[a,b[$	Half-open interval (open at the right)
(a,b), $]a,b[$	Open interval
$[a,\infty)$, $[a,\rightarrow[$	Interval closed at the left, infinite to the right
$(-\infty,\infty)$, $]\leftarrow,\rightarrow[$	Set of all real numbers
$\max_{x \epsilon X} f(x)$, $\max\{f(x)\|x \epsilon X\}$	Maximum of $f(x)$ when x is in the set X
min	Minimum
sup, l.u.b.	Supremum, least upper bound
inf, g.l.b.	Infimum, greatest lower bound
$\lim_{x\to a} f(x) = b$, $\lim_{x=a} f(x) = b$, $f(x) \to b$ as $x \to a$	b is the limit of $f(x)$ as x approaches a
$\lim_{x\to a-} f(x)$, $\lim_{x=a-0} f(x)$, $f(a-)$	Limit of $f(x)$ as x approaches a from the left
lim sup, $\overline{\lim}$	Limit superior
lim inf, $\underline{\lim}$	Limit inferior
l.i.m.	Limit in the mean
$z = x + iy = re^{i\theta}$, $\zeta = \xi + i\eta$, $w = u + iv = \rho e^{i\phi}$	Complex variables
$\bar{z}$, z^*	Complex conjugate
Re, $\Re$	Real part
Im, $\Im$	Imaginary part
arg	Argument
$\frac{\partial(u,v)}{\partial(x,y)}$, $\frac{D(u,v)}{D(x,y)}$	Jacobian, functional determinant
$\oint$, $\oint$	Line integrals
$\int_E f(x)\, d\mu(x)$	Integral (for example, Lebesgue integral) of function f over set E with respect to measure μ
$f(n) \sim \log n$ as $n \to \infty$	$f(n)/\log n$ approaches 1 as $n \to \infty$
$f(n) = O(\log n)$ as $n \to \infty$	$f(n)/\log n$ is bounded as $n \to \infty$
$f(n) = o(\log n)$	$f(n)/\log n$ approaches zero
$f(x) \nearrow b$, $f(x) \uparrow b$	$f(x)$ increases, approaching the limit b
$f(x) \downarrow b$, $f(x) \searrow b$	$f(x)$ decreases, approaching the limit b
a.e., p.p.	Almost everywhere
ess sup	Essential supremum
C^0, $C^0(X)$, $C(X)$	Space of continuous functions
C^k, $C^k[a,b]$	The class of functions having continuous kth derivative (on $[a,b]$)
C'	Same as C^1
Lip_α, Lip α	Lipschitz class of functions
L^p, L_p, $L^p[a,b]$	Space of functions having integrable absolute pth power (on $[a,b]$)
L'	Same as L^1
(C,α), $)C,p)$	Cesàro summability

See COMPLEX NUMBERS AND COMPLEX VARIABLES; NUMBER THEORY.

Special functions

$[x]$	The integral part of x
$\binom{n}{k}$, ${}^nC_k$, ${}_nC_k$	Binomial coefficient $n!/k!(n-k)!$
$\left(\frac{n}{p}\right)$	Legendre symbol
e^x, exp x, $e \uparrow x$	Exponential function
sinh x, cosh x, tanh x	Hyperbolic functions
sn x, cn x, dn x	Jacobi elliptic functions
$\wp(x)$	Weierstrass elliptic function
$\Gamma(x)$	Gamma function
$J_\nu(x)$	Bessel function
$\chi_X(x)$	Characteristic function of the set X: $\chi_X(x) = 1$ in case $x \epsilon X$, otherwise $\chi_X(x) = 0$
sgn x	Signum: sgn $0 = 0$, while sgn $x = x/\|x\|$ for $x \neq 0$
$\delta(x)$	Dirac delta function

Algebra, tensors, operators

$+$, $.$, $\times$, $\circ$, $\top$, τ	Laws of composition in algebraic systems
e, 0	Identity, unit, neutral element (of an additive system)
e, 1, I	Identity, unit, neutral element (of a general algebraic system)
e, ϵ, E, P	Idempotent
a^{-1}	Inverse of a
Hom(M,N)	Group of all homomorphisms of M into N
G/H	Factor group, group of cosets
$[K:k]$	Dimension of K over k
$\oplus$, $\dotplus$	Direct sum
$\otimes$	Tensor product, Kronecker product
$\wedge$	Exterior product, Grassmann product
$\vec{x}$, $\mathbf{x}$, $\mathfrak{x}$, x	Vector
$\vec{x} \cdot \vec{y}$, $\mathbf{x} \cdot \mathbf{y}$, $(\mathfrak{x},\mathfrak{y})$	Inner product, sca uct, dot produc

$\mathbf{x} \times \mathbf{y}$, $[\mathfrak{x},\mathfrak{y}]$, $\mathbf{x} \wedge \mathbf{y}$	Outer product, vector product, cross product
$\lvert x \rvert$, $\lvert x \rvert$, $\lVert x \rVert$, $\lVert x \rVert_p$	Norm of the vector x
Ax, xA	The image of x under the transformation A
δ_{ij}	Kronecker delta: $\delta_{ii} = 1$, while $\delta_{ij} = 0$ for $i \neq j$
A', ^tA, A^t, $'A$	Transpose of the matrix A
A^*, $\bar{A}$	Adjoint, Hermitian conjugate of A
tr A, Sp A	Trace of the matrix A
det A, $\lvert A \rvert$	Determinant of the matrix A
$\Delta^n f(x)$, $\Delta_h{}^n f$, $\underset{h}{\Delta}{}^n f(x)$	Finite differences
$[x_0,x_1]$, $[x_0,x_1,x_2]$, $\underset{x_1}{\Delta} u_{x_0}$, $[x_0,x_1]_f$	Divided differences
∇f, grad f	Read "gradient of f"
$\nabla \cdot \mathbf{v}$, div $\mathbf{v}$	Read "divergence of $\mathbf{v}$"
$\nabla \times \mathbf{v}$, curl $\mathbf{v}$, rot $\mathbf{v}$	Read "curl of $\mathbf{v}$"
∇^2, Δ, div grad	Laplacian
$[X,Y]$	Poisson bracket, or commutator, or Lie product
$GL(n,R)$	Full linear group of degree n over field R
$O(n,R)$	Full orthogonal group
$SO(n,R)$, $O^+(n,R)$	Special orthogonal group

Topology

E^n	Euclidean n space
S^n	n sphere
$\rho(p,q)$, $d(p,q)$	Metric, distance (between points p and q)
$\bar{X}$, X^-, cl X, X^c	Closure of the set X
FrX, fr$\dot{X}$, ∂X, bdry X	Frontier, boundary of X
int X, $\mathring{X}$	Interior of X
T_2 space	Hausdorff space
F_σ	Union of countably many closed sets
G_δ	Intersection of countably many open sets
dim X	Dimensionality, dimension of X
$\pi_1(X)$	Fundamental group of the space X
$\pi_n(X)$, $\pi_n(X,A)$	Homotopy groups
$H_n(X)$, $H_n(X,A,G)$, $H_*(X)$	Homology groups
$H^n(X)$, $H^n(X,A;G)$, $H^*(X)$	Cohomology groups

See Algebra; Calculus of tensors; Calculus of vectors; Operator theory.

Probability and statistics

X, Y	Random variables
$P(X \leq 2)$, Pr $\{X \leq 2\}$	Probability that $X \leq 2$
$P(X \leq 2 \mid Y \geq 1)$	Conditional probability
$E(X)$, $\mathcal{E}(X)$	Expectation of X
$E(X \mid Y \geq 1)$	Conditional expectation
c.d.f.	Cumulative distribution function
p.d.f.	Probability density function
c.f.	Characteristic function
$\bar{x}$	Mean (especially, sample mean)
σ, s.d.	Standard deviation
σ^2, Var, var	Variance
$\mu_1, \mu_2, \mu_3, \mu_i, \mu_{ij}$	Moments of a distribution
ρ	Coefficient of correlation
$\rho_{12 \cdot 34}$	Partial correlation coefficient

See Mathematics; Probability; Statistics.

[C.D.; H.A.P.]

Mathematical physics

The aim of mathematical physics is to deduce the consequences of the more established physical theories by relying mainly on the method of mathematical solution; the basic laws of physics are presumed known. A fruitful approach is possible largely because there is close analogy between the mathematical problems arising in different fields of theoretical physics. The same partial differential equations are encountered in many different contexts.

Some examples of problems in mathematical physics are the following:

1. The theory of the motion of planets, particularly the classical three-body problem; for example, the motion of an asteroid under the combined influence of the Sun and Jupiter. Gyroscopic motion of rigid bodies.

2. Potential theory, applicable primarily in electrostatics and hydrodynamics of nonviscous flow. Many important special functions such as Bessel functions and Legendre polynomials have been developed in connection with potential theory. Functions of a complex variable are useful for two-dimensional problems.

3. Theory of vibrations, determining the normal modes of electromagnetic or elastic vibration of regions of a given shape, or of systems of bodies interacting in various ways. Among other fields this is important for the theory of microwave cavities, for acoustics, and for seismology. Again special mathematical functions are important.

4. Wave propagation, including the exact solution of diffraction problems such as for electromagnetic or acoustic waves.

5. Solution of problems in wave mechanics, such as are encountered in the helium atom or the hydrogen molecule, which are too complicated for direct analytical solution and yet simple enough to be solved accurately.

6. Diffusion problems, such as diffusion of neutrons in matter, conduction of heat, and transport phenomena in statistical mechanics.

7. Nonlinear problems in hydrodynamics, elasticity theory, and so on.

8. Problems in probability theory related to statistical mechanics.

Through World War II, the main technique of mathematical physics was the analytical mathematical solution of problems. Since World War II, high-speed computing machines have become increasingly important and have made numerical solutions possible for many problems where the analytical technique did not work.

The term mathematical physics is sometimes used synonymously with theoretical physics, especially in Great Britain. *See* THEORETICAL PHYSICS. [H.A.BE.]

Mathematics

Mathematics is frequently encountered in association and interaction with astronomy, physics, and other branches of natural science, and it also has deep-rooted affinities to what are called humanities. Actually, it is a realm of knowledge entirely by itself, and one of considerable scope; the word mathematics stems from a root which means learnable knowledge. Mathematical knowledge is commonly deemed to have a high degree of validity, irrespective of man's cultural conditioning and predilection, although it can be argued that, in the past, cultural settings have affected its development noticeably.

Relation to science. As far as the scientist is concerned, mathematics is not a branch of natural science itself. It does not deal with phenomena and objects of the external world and their relations to each other, but, strictly speaking, only with objects and relations of its own imagery. Mathematical figures in two- or three-dimensional geometry are largely idealizations of objects occurring in the physical world, but figures in n-dimensional space for general n no longer are. Integer numbers 1, 2, 3, . . . , and even real numbers in general, can be claimed to be abstractions from quantities occurring in the physical world, but the "imaginary" number $i = \sqrt{-1}$ has received its name from the very fact that it has no basis in the material world, even though the use of complex numbers $a + bi$ is indispensable to science. If a laboratory experiment idealizes a physical system in order to eliminate secondary features not essential to the study at hand, its ultimate objective is an understanding of the unidealized physical system. But even if it were true that the five regular solids as investigated in the 13th book of Euclid's *Elements* (cube, tetrahedron, octahedron, icosahedron, and dodecahedron) had been found because of occurrence of approximating crystals in nature, nevertheless, once found, the idealized geometric figures would still have become primary objects of consideration in mathematics.

Mathematics is not subordinate to natural science by being a handmaiden of it, and one can practice competently meaningful mathematics without being concerned with science at all. Especially, philosophical attempts to reduce all origin of mathematics to utilitarian motives are wholly unconvincing. However, it is fair to say that mathematics is the language of science in a deep sense. Mathematics is an indispensable medium by which and within which science expresses, formulates, continues, and communicates itself. And just as the language of true literacy not only specifies and expresses thoughts and processes of thinking but also creates them in turn, so does mathematics not only specify, clarify, and make rigorously workable concepts and laws of science but also at certain crucial instances becomes an indispensable constituent of their creation and emergence as well. In Isaac Newton's formula for the motion of a particle on a straight line

$$m \frac{d^2x}{dt^2} = F$$

the mass m and the force F are nonmathematical objects, perhaps. But the instantaneous velocity $v = dx/dt$ and the instantaneous acceleration $a = dv/dt = d^2x/dt^2$ are wholly mathematical, and without a mathematical theory of the infinitesimal calculus, they are not conceivable. Newton, the physicist, was driven to creating his version of the calculus because of this. Also, Newton had to have not only the process of differentiation but also the concept of a mathematical function, because only a function can be differentiated. He required not only the path function $x = x(t)$, but, for the second derivative, he had to envisage the velocity $v = dx/dt$ itself again as a function depending on t, even though in the definition of v this dependence had been reduced to 'instantaneousness." This concept of function was given to Newton by the then new theory of analytic geometry of René Descartes, and it is a fact that after Archimedes' work on the motion of a lever, theoretical mechanics stood virtually still for almost 2000 years until the twin mathematical concepts of function and derivative were ready to emerge. *See* CALCULUS, DIFFERENTIAL AND INTEGRAL.

A striking case of the manner in which mathematics may shape physics has occurred in the theory of relativity. When Albert Einstein was pondering the transition from special to general relativity, his attention was drawn to G. Ricci's work on tensor calculus. Einstein found it an excellent medium for his thoughts, adopted it, and made it widely known to mathematicians. In the 1920s the quantum theory could move as rapidly as it did only because certain prefabricated parts from the theory of matrices and differential equations were lying about somewhat idly. *See* DIFFERENTIAL EQUATION; MATRIX THEORY.

Creative formulas. A formula is a string of mathematical symbols subject only to certain general rules of composition. To a working mathematician a string of symbols is a formula if it is something worth remembering. Much mathematics is concentrated in and propelled by certain formulas of unusual import.

The oldest such is the Pythagorean theorem $c^2 = a^2 + b^2$, where a, b are the sides and c is the hypotenuse in a right-angled triangle. If a, b are rational numbers p/q, then c need not be so; for instance, when $a = b = 1$, $c = \sqrt{2}$ and is not a rational number. This compelled the Greeks to study ratios of incommensurable quantities and also quadratic irrationalities. In modern times the formula was extended to $c^2 = a_1^2 + \cdots + a_n^2$ [illegible] of a rectangular parallel-

n-dimensional euclidean space, and gave the expression

$$s^2 = (x^1 - y^1)^2 + \cdots + (x^n - y^n)^2 \quad (1)$$

for the distance between two points $x = (x^1, \ldots, x^n)$, $y = (y^1, \ldots, y^n)$ in cartesian coordinates there. For infinitesimally near points this becomes the line element $ds^2 = (dx^1)^2 + \cdots + (dx^n)^2$ which in the more general affine version

$$ds^2 = \sum_{i,j=1}^{n} g_{ij}\, dx^i\, dx^j \qquad (2)$$

is the cornerstone of Riemannian geometry On the other hand, analytic developments demanded the extension of Eq. (1) to a space of infinitely many dimensions and then even to the case in which coordinates are complex numbers; thus

$$s^2 = |z^1 - w^1|^2 + |z^2 - w^2|^2 + \cdots$$

With this definition of distance the space becomes a Hilbert space, the basis of operator theory and quantum mechanics. Finally, again for finite dimension but still complex coordinates, it became significant, in analogy to Eq. (2), to introduce the line element

$$ds^2 = \sum_{\alpha,\beta=1}^{n} g_{\alpha\bar\beta}\, dz^\alpha\, d\bar z^{\beta} \qquad (2')$$

It is the basis for the study of Riemannian geometry on complex manifolds, and the study of such manifolds is penetrating even into relativity and field theory. *See* GEOMETRY, RIEMANNIAN; OPERATOR THEORY; QUANTUM MECHANICS

The simple relation $(1 + x)^2 = 1 + 2x + x^2$ was generalized by Newton to

$$(1 + x)^n = 1 + \binom{n}{1} x + \binom{n}{2} x^2 + \cdot \qquad (3)$$

The coefficients are the binomial coefficients

$$\binom{n}{k} = \frac{n(n-1) \quad (n-k+1)}{1 \quad 2 \quad\quad k} \qquad (4)$$

and the formula is a generating relation for them. The expression in Eq (4) can also be set up if n is any nonintegral number, real or complex, and it was indeed established that Eq. (3) holds for any n for $|x| < 1$. The expansion is usually an *infinite* series, and it is in fact the Taylor series of $(1 + x)^n$. By studying, for general n, the behavior of the series as $|x| \to 1$, H. Abel laid the foundation for the theories of summability of nonconvergent series. *See* SERIES.

Euclid's theorem that any integer is a unique product of primes (stated in Euclid only *implicitly*) can be expressed as a formula

$$a = p_1^{n_1} \times p_2^{n_2} \times \cdots \times p_k^{n_k}$$

The concept of a prime factor and the quest for a theorem on unique decomposition *into prime factors* has been a leading principle of *arithmetic* and algebra on all levels ever since. *See* NUMBER THEORY.

The constant π, widely known from antiquity through the formulas $l = 2\pi r$, $A = \pi r^2$ for a circle and $A = 4\pi r^2$, $V = 4/3\pi r^3$ for a sphere, also enters the sophisticated formula

$$\frac{\pi}{4} = 1 - \tfrac{1}{3} + \tfrac{1}{5} - \tfrac{1}{7} + \cdots$$

derived by G. W. Leibnitz. Both kinds of formulas were much used for the approximate computation of this transcendental number, and a great deal of computational technique (including the use of computing machines) was initiated in the course.

The transcendental number e, the basis of natural logarithms, appears in the eighteenth century in the formulas

$$e^{2\pi\sqrt{-1}} = 1 \qquad \cos\varphi + \sqrt{-1}\sin\varphi = e^{\varphi\sqrt{-1}}$$

which have been of immeasurable consequence to all of mathematics. Even the age-old plane trigonometry was greatly streamlined by them. *See* NUMERICAL ANALYSIS.

Near the end of the eighteenth century, L. Euler gave for convex polyhedrons the formula $e_0 -$ [illegible] number of vertices. [illegible] and [illegible] in algebraic geometry, and it was a starting point for H. Poincaré's combinatorial topology. Poincaré's own formula $B_p = B_{n-p}$ for Betti numbers of compact manifolds is even today one of the most arresting statements of the theory.

Cauchy's formula

$$f(z) = \frac{1}{2\pi i} \oint_C \frac{f(\zeta)\, d\zeta}{\zeta - z}$$

is a centerpoint from which the theory of complex functions radiates. *See* COMPLEX NUMBERS AND COMPLEX VARIABLES.

Galileo's *main achievement* is embodied in the formula $s = \tfrac{1}{2}gt^2$ for a falling body, and *Newton's* gravitational law in the formula $F = m_1m_2/r^2$ which also represents Coulomb's law of electricity. Albert Einstein's daring formula $E = mc^2$ from special relativity created the atomic bomb, and W. Heisenberg's interpretation (1927) of the formula $pq - qp = h/2\pi i$ from quantum theory as an indeterminacy relation for physical observables has been an irritant to philosophers ever since. *See* RELATIVITY; UNCERTAINTY PRINCIPLE.

Foundations—mathematical logic. A prime demand on mathematics is that it be deductively rigorous, and a traditional model for intended rigor is Euclid's presentation of mathematical assertions in theorems. For the philosopher G. Spinoza in the seventeenth century, "more geometrico" is a synonym for "deductively rigorous." A theorem is a proposition which has been proved, excepting certain first theorems called axioms which are admitted without proof; and to prove a theorem means to obtain it from other theorems by certain procedures of deduction or inference. It had been a commonplace for long that each branch of mathematics is based on its own axioms, but during the

nineteenth century, mathematicians arrived at the insight that even the same branch might have alternate axioms. Specifically, there were envisaged alternate versions of two- and three-dimensional geometry, the axiom varied being the axiom on parallels. It was also recognized that a set of axioms becomes mathematically possible if it is logically consistent, that is, if one cannot deduce from the axioms two theorems one of which, as a proposition, is the negation of the other. *See* GEOMETRY, EUCLIDEAN; GEOMETRY, NONEUCLIDEAN.

At the same time certain developments led to the realization that not only the axioms but the rules of inference themselves might be, and even ought to be, subject to variations. Now if axioms and rules of inferences are both viewed as subject to change, it is customary to speak of a mathematical system or also a formal system, and, of course, an irreducible first requirement is that the system shall be consistent after the manner just stated. Consistency alone is a somewhat negative property. There is a further property, called completeness, which is more positive, and which, if present, is very welcome. A system is complete if for any proposition which can be formulated it either can be proved that it holds or that its negation holds If a theorem holds in a system and if the system is altered, then the same proposition, or what corresponds to it in the new system, may become doubtful if not outright false; and if it remains valid it may require a new proof because certain axioms or rules of inference are no longer available

Some of the developments which led to doubt as to whether the traditional rules of inference are inviolate were the following:

1. G. Boole had found (1854) that the classical Aristotelian connectives "and," "or," "negation of" for propositions follow rules similar to those which the operations addition, multiplication, "the negative of" obey in ordinary algebra (Boole's algebra of propositions) and this gradually took from rules of inference the privileged status of untouchability.

2. G. Cantor, the founder of the theory of sets and operations between them, defined a set (intuitively or naively) as the collection of all objects having a certain property which is verbally expressible. Especially "the set of all sets" is again

paradox). Divide the totality of all possible sets into two categories. A set shall belong to category I if it does not contain itself as an element, and to category II if it does contain itself as an element. Now form the set M whose elements are the sets of category I. It can now be reasoned by deductive steps admissible in Cantor's own theory that the set M cannot belong to either of the two categories, although the original division into categories did assign each set to one of them.

3. In 1904, E. Zermelo formulated the following axiom of choice. Given any family of nonvacuous sets $\{S\}$, no matter how (infinitely) large the family may be, it is possible to choose simultaneously an element $x = x_S$ from each given set S and thus to consider the set M consisting of precisely these elements. By the use of this axiom, some striking theorems in classical mathematics could be proven, which, without the use of the axiom, seemed to be logically out of reach entirely. Mathematicians began to wonder whether a theorem based on the axiom of choice is indeed valid or, at any rate, whether it has the same level of validity as one without it, and as a consequence of this, theorems employing the axiom of choice were being frequently labeled as such. *See* SET THEORY.

Some of the doubts were resolved eventually; the most striking results to date are the following ones of K. Gödel. (1) Any consistent mathematical system which is sufficient for classical arithmetic must be incomplete. In other words, in a formal system which is expressible within the arithmetical language taught to a student of a secondary school in any country it is possible to formulate assertions which can be neither proved nor disproved. (2) Any such system remains consistent if one adds to it the axiom of choice, and it also remains consistent if one adds the negation of the axiom of choice. In other words, the system of working mathematics can neither disprove nor prove the axiom of choice. (3) The so-called general continuum hypothesis is also consistent with ordinary mathematics, and, in fact, ordinary mathematics remains consistent if the axiom of choice and the general continuum hypothesis are added simultaneously.

G. Cantor says that two sets A_1, A_2 have the same cardinal number if there exists a one-to-one correspondence between the elements of A_1 and A_2. Any infinite set has the same cardinal number as some proper subset of itself. For instance, the set of all integers $A_1 = \{1,2,3, \ldots\}$ has the same cardinal number as the subset of all even integers $A_2 = \{2,4,6,8, \ldots\}$ because the association of the number n in A_1 to the number $2n$ in A_2 is one-to-one. However, for a finite set this cannot happen, and the cardinal number of a finite set is the ordinary number of its elements. Next, Cantor says that a set A_2 has a larger cardinal number than A_1 if there is a one-to-one correspondence from A_1 to a subset of A_2 but not one from A_2 to A_1 or a subset of A_1. He then shows that if A is any set, and S is the set of its subsets including the empty set, the cardinal number of S is greater than that of A. If A is finite and has n elements, then S has 2^n elements. Now 2^n is indeed greater than n, and for $n \geqq 2$ there are even other integer numbers between n and 2^n. This is true for finite sets. However, for infinite sets the situation seems to change radically, and the generalized continuum hypothesis asserts that there are no other cardinal numbers between those of A and S. *See* LOGIC.

Counterexamples. It follows easily from the definition of a derivative that a differentiable function $y = f(x)$ in $a < x < b$ is also continuous. K. Weierstrass deepened considerably this statement by demonstrating that the converse is tr nstructed a function $f(x)$ which is

tinuous in $a < x < b$ but does not have a derivative at any point (but his function does have right and left derivatives at all points). A construction of this kind is called a counterexample. Systematic preoccupation with counterexamples, which sometimes require considerable ingenuity, did not arise before the middle of the nineteenth century. Another early counterexample of considerable consequence referred to Fourier series. If $f(x)$ is periodic with period 2π then its Fourier series is

$$\tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \quad (5)$$

where, by definition,

$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos nt\,dt \quad b_n = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin nt \quad (6)$$

Fourier himself stated somewhat vaguely *that for "any" function $f(x)$* the partial sums of the Fourier series are convergent to $f(x)$. P. G. L. Dirichlet was the first to give a specific criterion of convergence. In his criterion the function $f(x)$ *may have* certain simple discontinuities. But *in the present* discussion these discontinuities *are excluded* and $f(x)$ must be continuous, first of all. Dirichlet proved that the Fourier series converges everywhere to $f(x)$, if $f(x)$ is piecewise *monotone*, that is, if the interval of periodicity $0 \leqq x < 2\pi$ may be divided into a *finite number* of smaller intervals such that *in each of these* the function is either monotonely *increasing* or monotonely decreasing. Then, P. du Bois Reymond demonstrated *that this* requirement of monotoneity *cannot be dispensed* with entirely. He constructed a *continuous periodic* function $f(x)$ for *which the* Fourier series fails to converge in at least one point. By a general principle developed by A. Harnack (condensation of singularities) it is then possible to construct *a continuous* function whose Fourier series *fails* to converge at countably many points *even if* they are everywhere dense, *for instance*, at all rational points $x = p/q$. But *to this day it* is not known whether there exists a continuous function whose Fourier series fails to converge at all *points*, or at least at all points with the only *exception* of a set of Lebesgue measure 0. *See* FOURIER SERIES AND INTEGRALS.

Counterexamples frequently *fall into* patterns, meaning that the same underlying construction is used over again for different but related *purposes*. A construction used in many counterexamples is Cantor's ternary set. Divide the closed *interval* $0 \leqq x \leqq 1$ into three equal parts $(0, \frac{1}{3})$, $(\frac{1}{3}, \frac{2}{3})$, $(\frac{2}{3}, 1)$ and remove the middle one, but *only* the open inner part of it $\frac{1}{3} < x < \frac{2}{3}$. This leaves two closed intervals $0 \leqq x \leqq \frac{1}{3}$, $\frac{2}{3} \leqq x \leqq 1$. In *each of these* remove the open inner third, that is, $\frac{1}{9} < x < \frac{2}{9}$ and $\frac{2}{3} + \frac{1}{9} < x < \frac{2}{3} + \frac{2}{9}$. This *leaves four* closed intervals altogether. From each, remove the open inner third, and so on. *After n steps* the number of intervals removed is $1 + 2 + 2^2 + \cdots + 2^{n-1}$ *and* the sum of the length of the removed intervals is

$$\frac{1}{3} + \frac{2}{3^2} + \frac{2^2}{3^3} + \cdots + \frac{2^{n-1}}{3^n} = 1 - \left(\frac{2}{3}\right)^n$$

This value is less than 1, which must be so, because the total interval has length 1 and the removed intervals do not overlap. However if one lets n go to ∞, then $(\frac{2}{3})^n$ tends to 0, and the aggregate length of all infinitely many intervals which have been removed has value 1. Thus, the pointset which is left over ought somehow to have a one-dimensional measure which has value 0. This is so indeed in the theory of Lebesgue. *Also the* left-over set is nowhere dense. However, it is a very big set still, by a crude counting of its points, at any rate. More precisely, it turns out that its cardinal number is the same as for the entire interval $0 \leqq x \leqq 1$. This can be deduced from the following fact which is interesting for many purposes. Any number $0 \leqq x \leqq 1$ can be represented, instead of by a decimal expansion, by a ternary expansion, that is, by an expansion $0.a_1a_2a_3 \ldots$ in which each symbol a_1, a_2, a_3, $\ldots$, has one of the values 0, 1, 2. Now, the left-over set consists of precisely those points in whose expansion the symbols a_1, a_2, a_3, $\ldots$, assume only the values 0, 2 and not 1.

Constructiveness and approximations. Some mathematicians object to mere existence proofs, and they demand that any proof shall also be constructive. The interpretations of this demand differ widely. Some proofs closely approach what a practical mathematician welcomes; if, for instance, a theorem asserts the existence of a number or a function, then the proof must also embody a procedure for actual computation of the solution, approximately, at least. Other versions are little more than the negative requirement that certain *combinations* of inference shall be avoided. *There are* also views which combine both; the best *known* among the last is the intuitionist view. It firmly demands a certain kind of constructiveness, which, however, does not necessarily guarantee the calculation by present-day computing machines. However, the actual stricture by which intuitionism became widely known is that proof by contradiction is not admissible. It is also called proof by double negation, and it is equivalent to the Aristotelian *tertium non datur*. It assumes tentatively that the proposition to be *proved* is false and from this assumption deduces a contradiction to a previously established theorem.

There is one difference between practical and theoretical demands on constructiveness. Sometimes it is possible to prove the existence of a function nonconstructively *and after* that, based on the existence statement, *to devise a procedure* for a manageable construction. To practical mathematics this is satisfactory; to theoretical, it is not.

A demand superficially related to constructiveness, and without any overtones of "logic," is a desire to estimate the degree of approximation, whenever a task of approximation is being performed. For instance, a *classical result of* J. L. Lagrange for the Taylor series with remainder term states

that if $f(x)$ has n derivatives in the interval $0 \leqq x \leqq 1$, then $f(x) = P_n(x) + R_n(x)$ where $P_n(x)$ is the polynomial

$$P_n(x) = \sum_{m=0}^{n} \frac{f^{(m)}(0)}{m!} x^m \qquad (7)$$

and the "error term" $R_n(x)$ is in numerical value bounded by $M_n/n!$, where M_n is the maximum of $|f^{(n)}(x)|$ in $0 \leqq x \leqq 1$. This result leads to the following topic. The polynomial (7) can be formed only if $f(x)$ has n derivatives. However, there is a theorem of Weierstrass that any function $f(x)$ *which is continuous in an interval* $a \leqq x \leqq b$ *(and* nothing more) can be approximated uniformly by polynomials

$$P_n(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n \qquad (8)$$

Each proof for this theorem produces its polynomials, and there is an extensive theory for securing those for which the approximation is a best possible one.

The parallel question of approximating not by ordinary polynomials (8) but by trigonometric polynomials

$$s_n(x) = \tfrac{1}{2}a_0 + \sum_{m=1}^{n} (a_m \cos mx + b_m \sin mx) \qquad (9)$$

leads to mathematical procedures which at present pervade all of science. If one wishes to make the difference

$$\mathrm{Max}_{0 \leq x < 2\pi} |f(x) - s_n(x)| \qquad (10)$$

rather small, then the partial sums of the Fourier series (5) are not too good, meaning that it is not very advantageous to use the polynomials (9) in which the coefficients a_m, b_m are the expressions (6). However, it improves the degree of approximation if one replaces the partial sums by their arithmetic means (Fejer sums)

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_{n-1}}{n}$$

which are again exponential polynomials, namely

$$\sigma_n(x) = \tfrac{1}{2}a_0 + \sum_{m=1}^{n} \left(1 - \frac{m}{n}\right)(a_m \cos mx + b_m \sin mx)$$

There are other averages which approximate even better, and one is led to studying procedures of averaging and smoothing which are of importance to analysis, probability, and statistics. But a major development begins if, following F. W. Bessel and M. A. Parseval (nineteenth century), the expression (10) is replaced by the expression

$$\left[\frac{1}{2\pi}\int_0^{2\pi} |f(x) - s_n(x)|^2 \, dx\right]^{1/2} \qquad (11)$$

as a measure for the degree of approximation. In this case the partial sums of the Fourier series are indeed the best approximating sums for any order n, not only for ordinary Fourier series but for so-called orthogonal series in general. Pertinent application to the theory of accoustical and then optical waves led to an interpretation of the Fourier series (5) as a spectral resolution of its function $f(x)$, and this interpretation has been gradually extended to a host of related and analogous expansions in analysis, algebra, theory of probability, quantum theory, and other parts of physics. The measuring of approximation, deviation, and dispersion by a square of integral mean—which is Eq. (11)—has become a strong and pervasive influence on scientific thinking.

Space in mathematics. If geometry is the mathematics of space, then, in a superficial sense, all mathematics began with geometry, because apparently it began with measurements of figures: length, area, volume, size of angles. It did not concern itself with questions of shape but with clarifying and deciding when figures are equal or substantially equal with regard to form. The first true theory of geometry was the great theory of the Greeks, whose primary concern was study of the basic concept of equality of figures—their congruence and similarity—and the Greeks were so determined to dissociate their theory from the preceding phase of merely making measurements that Euclid's extensive work, for instance, avoids to a fault any kind of actual measurements. In the case of the Pythagorean theorem $c^2 = a^2 + b^2$, there is no hint that it might be an equality between numbers. It is only an equality between areas based on congruences, two of the squares being cut up into pieces and then put together into the third square after the manner of a jigsaw puzzle. But, for all its lofty purposes, Greek geometry was too rigid and circumscribed to be able really to cope with the mathematical problem of space. It came to a standstill and geometry did not progress further until, with the advent of coordinate systems, introduced by Descartes and his predecessors, a better mathematics of space could be initiated.

If a cartesian coordinate system is etched into two- or three-dimensional euclidean space, then the space becomes a pointset, each point being a pair (x^1,x^2) or a triple (x^1,x^2,x^3) of real numbers, and any figure a suitable subset of it. This is a deliberate process of arithmetization of space which unifies space and number at the base. This does not hamper geometry in its task of pursuing problems of shape but instead aids it. In the cartesian plane, two figures are similar if the points of one can be obtained from the points of the other by means of a transformation

$$\begin{aligned} y^1 &= a^1 + \alpha_1{}^1 x^1 + \alpha_2{}^1 x^2 \\ y^2 &= a^2 + \alpha_1{}^2 x^1 + \alpha_2{}^2 x^2 \end{aligned} \qquad (12)$$

where

$$\alpha_1{}^1 \cdot \alpha_1{}^2 + \alpha_2{}^1 \cdot \alpha_2{}^2 = 0 \qquad (13)$$

and for some $\rho > 0$

$$(\alpha_1{}^1)^2 + (\alpha_2{}^1)^2 = (\alpha_1{}^2)^2 + (\alpha_2{}^2)^2 = \rho^2 \quad (14)$$

The similarity is a congruence if, and only if, $\rho = 1$ (orthogonal transformation). Now this analytic

representation of congruence and similarity suggests a geometric examination of the most general linear transformations (12) which are nonsingular, that is, for which the determinant $|a_q^p| \neq 0$. They were virtually unknown to the Greeks, *although* they highlight the axiom of parallels of Euclid's geometry. A one-to-one transformation of the cartesian plane is such a linear transformation if, and only if, it carries a straight line into a straight line and parallel straight lines into parallel straight lines. Thus, a parallelogram goes into a parallelogram, and, in fact, given any two parallelograms there is a linear transformation which carries the one into the other no matter what the angles and ratios of sides in the two figures are. There is a geometry, the so-called affine geometry, in which any two parallelograms are considered equal. It cannot measure angles, and nonparallel segments cannot be compared for length. This geometry does have conics, however, and it can separate them into ellipses, parabolas, and hyperbolas. *See* CONFORMAL MAPPING.

The family of all linear transformations constitutes a transitive group and the subfamily of orthogonal transformations is already a transitive group. F Klein has made the pronouncement, which is generally accepted, that there arises a geometry on a space if on the space there is given a transitive group of transformations; two figures are considered equal whenever one figure can be carried into the other figure by one of the transformations. For the noneuclidean geometries of Bolyai-Lobachevski-Gauss various models have been exhibited which conform to this view and perhaps the most interesting one is the following. In the plane (x^1,x^2) introduce the complex variable $z = x^1 + ix^2$ and consider the family of transformations

$$w = e^{i\theta}\,\frac{z - a}{1 - \bar{a}z}$$

for all constant complex numbers a for which $|a| < 1$ and all real numbers θ. They give rise to the following noneuclidean geometry. The space is not the entire z plane but only the unit disk: $|z| < 1$. A "point" is an ordinary point in it and a "straight line" of the new geometry is in the euclidean geometry either a diameter of the disk or any circular arc inside the disk whose endpoints are on the boundary line of the disk and which meets the boundary line of the disk at right angles. Through a point outside a straight line of the geometry there go infinitely many straight lines which are "parallel" to it; that is, they do not meet it inside the disk. This makes the geometry a hyperbolic one. If, however, in a geometry any two straight lines always intersect, the geometry is called elliptic. The prototype of an elliptic geometry is a geometry on a surface of a sphere (for example, the surface of the earth) in which the straight lines are the great circles. As a matter of fact, the hyperbolic geometry just described is in a very precise sense a counterpart to the elliptic geometry of great circles on a sphere.

The arithmetization of space led to a purely mathematical creation of n-dimensional space, euclidean and other, for any integer dimension n, by defining its points generally as n-tuples of real numbers $(x^1, \ldots, x^n)$ with suitable definitions for various *geometrical* relations between such points. The most eye-catching consequence in science was the four-dimensional space of the theory of relativity, *but actually multidimensional* geometry had been playing a part in physics before that. If a mechanical system involves M mass points, it was customary in effect to *introduce the* space of dimension $n = 3M$ whose points are the states of the system, that is, the n-tuples of *coordinates* $\{x_m^1, x_m^2, x_m^3\}$, $m = 1, \ldots, M$, at any one time point. Also, if there are restraints operative in the system, then the Lagrange-Hamilton theory *suitably* reduced the dimension of the space by the use of the free parameters of the system instead of the *original* n coordinates themselves. The use of free parameters spread from mechanical systems to other systems in physics and chemistry, and all so-called equations of state are geared to such. *Finally, in* quantum theory a state of a system has infinitely many coordinates and the infinitely *dimensional* space representing it is a Hilbert space. Also, partly under the influence of Hilbert space, mathematicians have become fascinated with infinitely *dimensional* spaces in general. They are being studied intensively, and large parts of mathematics are being pressed into these new frames of reference.

The arithmetization of space also is reflected in the ever-widening use of graphs and charts. Any tabulated dependence of a number y on a number x is a function $y = f(x)$ and hence representable as a curve on graph paper; a great *amount* of information is thus illustrated and stored. This is nothing else but a practical application of the concept of a manifold in modern topology *and* differential geometry, and, as frequently, the abstract formulation of a mathematical concept is but a circumlocution of what common sense dictates. *See* CALCULUS OF VECTORS; COORDINATE SYSTEMS, GRAPHICAL; *see also* ALGEBRA; ANALYTIC GEOMETRY; PROBABILITY; STATISTICS; TOPOLOGY. [S B]

Bibliography: N. J. Achieser, *Theory of Approximation*, 1956; A. Ambrose and M. Lazerowitz, *Fundamentals of Symbolic Logic*, 1948; A. A. Fraenkel, *Abstract Set Theory*, 1953; D. Jackson, *The Theory of Approximation*, 1930; P. C. Rosenbloom, *The Elements of Mathematical Logic*, 1951; P. Suppes, *Introduction to Logic*, 1957.

Matrix calculus

The derivative of a matrix $A(t)$ whose elements $a_{ij}(t)$ are functions of a variable t is defined as

$$\frac{dA}{dt} = \lim_{\Delta t \to 0} \frac{A(t + \Delta t) - A(t)}{\Delta t} = \left(\frac{da_{ij}}{dt}\right)$$

Thus dA/dt is formed by replacing its elements by their derivatives.

If the matrices A, B are functions of t,

$$\frac{d}{dt}(A+B)=\frac{dA}{dt}+\frac{dB}{dt} \qquad \frac{d}{dt}(AB)=\frac{dA}{dt}B+A\frac{dB}{dt}$$

In differentiating a product the order of the factors must be preserved. Thus

$$\frac{d}{dt}A^2=\frac{dA}{dt}A+A\frac{dA}{dt} \quad \text{not} \quad 2A\frac{dA}{dt}$$

From $A^{-1}A=I$ one finds

$$\frac{dA^{-1}}{dt}=-A^{-1}\frac{dA}{dt}A^{-1}$$

The integral of $A(t)$ is defined as the matrix whose elements are integrals of $a_{ij}(t)$.

Every square n by n matrix satisfies a polynomial equation of lowest degree, its minimum equation:

$$A^m+a_1A^{m-1}+\cdots+a_mI=0 \qquad m\leqq n$$

This equation may be used to express all powers of $A>m-1$ in terms of $I, A, A^2, \ldots, A^{m-1}$. Therefore any matric polynomial $f(A)$ of degree $k>m-1$ may be replaced by a polynomial $F(A)$ of degree $\leqq m-1$.

If $f(t)=\sum_{n=0}^{\infty}c_nt^n$ is a power series whose radius of convergence is r, the matric series

$$f(A)=c_0I+\sum_{n=1}^{\infty}c_nA^n$$

will converge if all characteristic roots of A are less than r in absolute value. The minimum equation of A may again be used to replace the power series $f(A)$ by a polynomial $F(A)$ of degree $\leqq m-1$.

If the minimum equation of A has distinct roots $\lambda_1, \lambda_2, \ldots, \lambda_m$, the polynomial $F(A)$ has the characteristic roots $F(\lambda_i)=f(\lambda_i)$. The scalar polynomial $F(t)$ is completely determined by the m known values $f(\lambda_i)$ it assumes when $t=\lambda_i$. Hence $F(t)$, and also $F(A)$, may be determined by Lagrange's interpolation formula:

$$F(A)=f(\lambda_1)L_1(A)+f(\lambda_2)L_2(A)+\cdots+f(\lambda_m)L_m(A)$$

$$L_i(A)=\prod_{i\neq j}(A-\lambda_jI)/\prod_{i\neq j}(\lambda_i-\lambda_j)$$

where each product in $L_i(A)$ has $m-1$ factors.

Differential equations. Let $x_1(t), \ldots, x_n(t)$ be n unknown functions and $X(t)$ their column vector. Then if $A(t)$ is an n by n matrix whose elements are continuous functions of t, the system of n linear differential equations

$$dX/dt=AX \qquad X(0)=X_0$$

may be solved by successive approximations. The nth approximation is obtained from the $(n-1)$th by integrating

$$dX_m/dt=AX_{m-1} \qquad (n=1,2,\ldots)$$

from $t=0$ to t. Then a solution is obtained in the form

$$X=\lim_{n\to\infty}X_n=\Omega X_0$$

$$\text{where} \quad \Omega=I+\int_0^t A\,dt+\int_0^t A\,dt\int_0^t A\,dt+\cdots$$

is called the matrizant of A. When A has constant elements, Ω reduces to

$$\Omega=I+At+\frac{1}{2!}A^2t^2+\frac{1}{3!}A^3t^3+\cdots=e^{tA}$$

Jacobian matrices. If $u=f(x,y)$, $v=g(x,y)$, the jacobian $\partial(u,v)/\partial(x,y)$ is the determinant of the jacobian matrix $\begin{pmatrix}u_x & u_y\\ v_x & v_y\end{pmatrix}$ whose elements are partial derivatives. If $x=\phi(s,t)$, $y=\psi(s,t)$,

$$\begin{pmatrix}u_x & u_y\\ v_x & v_y\end{pmatrix}\begin{pmatrix}x_s & x_t\\ y_s & y_t\end{pmatrix}=\begin{pmatrix}u_s & u_t\\ v_s & v_t\end{pmatrix}$$

by the chain rule. In particular if $s=u$, $t=v$ this becomes

$$\begin{pmatrix}u_x & u_y\\ v_x & v_y\end{pmatrix}\begin{pmatrix}x_u & x_v\\ y_u & y_v\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

These results admit obvious generalization to n by n jacobian matrices. The last equation shows that the matrices on the left are reciprocals; thus when one is known, its reciprocal gives all the derivatives in the other. See MATRIX THEORY. [L.BR.]

Bibliography: L. Brand, *Advanced Calculus*, 1955; R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices*, 1955

Matrix mechanics

A formulation of quantum theory in which the operators are represented by time-dependent matrices. For a discussion of quantum-mechanical operators and other information essential to an understanding of the present article, see QUANTUM THEORY, NONRELATIVISTIC.

Matrix mechanics is disadvantageous for obtaining quantitative solutions to actual problems; because it is concisely expressed in a form independent of special coordinate systems, however, matrix mechanics is advantageous for proving general theorems. For the purposes of the following brief discussion it is sufficient to consider a one-dimensional spinless system, described by a wave function $\psi(x,t)$.

A matrix A is a rectangular array of numbers; the number of rows or columns of the array may be finite or infinite; A_{mn} is the element (number) in the mth row and the nth column; m, n may be discrete or continuous indices (see MATRIX THEORY). Symbolic addition, $C=A+B$, is defined as meaning that $C_{mn}=A_{mn}+B_{mn}$ for each m and n, with the implication that the rows of A, B have the same indexing m, and similarly for the columns. Symbolic multiplication $C=AB$ is defined to mean $C_{mn}=\Sigma_jA_{mj}B_{jn}$, summed over all values of j, again with the implication that the columns of A

and the rows of B have the same indexing j, when j is continuous, the sum in the definition of C_{mn} is replaced by an integral. The adjoint $A^\dagger$ of A is obtained by interchanging rows and columns of A, and then taking the complex conjugate of each element, that is, $A_{mn}^\dagger = (A_{nm})^* \equiv A_{nm}^*$; $(AB)^\dagger = B^\dagger A^\dagger$.

Suppose $\psi(x,t)$ is expanded in terms of any complete set of orthonormal functions $u_n(x)$:

$$\psi(x,t) = \sum_n \psi_n(t) u_n(x) \qquad (1)$$

The matrix elements (in this u representation) of any quantum-mechanical operator A are defined by

$$A_{mn} = \int_{-\infty}^{\infty} dx\, u_m^*(x) A u_n(x) \qquad (2)$$

For instance, when $A \equiv p_x$, A is replaced by $(\hbar/i)\,\partial/\partial x$ in Eq. (2). From Eqs. (1) and (2) and the orthonormality of $u_n(x)$, if

$$A\psi = \phi = \sum_n \phi_n u_n$$

then

$$\phi_m = \sum_n A_{mn}\psi_n$$

This result shows that the equation $\phi = A\psi$, which states that ϕ is the function resulting from the operation A on ψ, equally well *can be interpreted* as a matrix equation, provided ψ, ϕ are represented by single-column matrices whose elements are respectively ψ_n, ϕ_n. Similarly, the expectation value of the operator A in the state ψ is

$$\langle A \rangle = \int \psi^* A \psi = \int \psi^* \phi = \sum_{m,n} \psi_m^* A_{mn} \psi_n \qquad (3)$$

that is, in matrix notation $\langle A \rangle = \psi^\dagger A \psi$; the adjoint of a column matrix ψ_m is the row ψ_m^*.

The time-dependent Schrödinger equation implies $\psi(t) = \exp(-iHt/\hbar)\psi(0)$. Thus, noting that because the operator H is Hermitian the matrix H is self-adjoint, $H^\dagger = H$, the expectation value of A at time t is

$$\langle A \rangle_t = \psi^\dagger(0) \exp(iHt/\hbar) A \exp - (iHt/\hbar)\psi(0) \qquad (4)$$

Eq. (4) shows that $\langle A \rangle_t$ can be computed in two equivalent ways: (i) the *conventional* Schrödinger representation, in which ψ is time-dependent and A_{mn} *are time-independent*; (ii) the Heisenberg *representation*, in which ψ is time-independent and equal to $\psi(0)$, but A is represented by time-dependent matrix elements

$$A_{mn}(t) = [\exp(iHt/\hbar) A(0) \exp - (iHt/\hbar)]_{mn} \qquad (5)$$

Differentiating Eq. (5),

$$\frac{d}{dt}A(t) = \frac{iH}{\hbar}\exp(iHt/\hbar)A(0)\exp - (iHt/\hbar)$$
$$- \exp(iHt/\hbar)A(0)\exp - (iHt/\hbar)\frac{iH}{\hbar}$$
$$= \frac{1}{i\hbar}[A(t)H - HA(t)] \qquad (6)$$

that is, $dA/dt = (i\hbar)^{-1}(A,H)$, where $(A,B) = AB - AB$.

It can be proved that (i) operators can consistently be replaced by equivalent matrices, defined by Eq. (2); and (ii) the Heisenberg time-dependent matrix formulation of quantum theory is completely equivalent to the Schrödinger time-dependent wave function formulation. [E.C.]

Matrix theory

A matrix is a rectangular array of numbers, or other elements, of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \qquad (1)$$

The above array A is an m by n matrix with m rows and n columns, and the size of A is said to be m by n. The rows of a matrix are always numbered from the top down and the columns from left to right. The position of each element in the array is given by its subscripts; that is, a_{ij} is the element in the ith row and jth column. Since every element of A is represented by a_{ij} as i takes on the values $1, 2, \ldots, m$ and j the values $1, 2, \ldots, n$, a_{ij} is called the typical element of A, and the compact notation $A = (a_{ij})$ is used when the size of A is given.

Matrices have application as computational devices in such widely diversified fields as economics, psychology, statistics, engineering, physics, and mathematics. In mathematics, matrices are useful tools in the study of linear systems of algebraic equations, linear differential equations, linear mappings and transformations, and bilinear and quadratic forms.

For example, in the linear system of equations

$$\begin{aligned} -3x + 2y - 6z &= 10 \\ 7x - y + 3z &= 0 \\ x + y - 5z &= 1 \end{aligned}$$

the letters x, y, and z are merely symbols which stand for possible numerical solutions, and the only significant features of this system are the numbers which appear in the equations and their relative positions. Therefore, these equations are completely described by the 3 by 4 matrix

$$\begin{pmatrix} -3 & 2 & -6 & 10 \\ 7 & -1 & 3 & 0 \\ 1 & 1 & -5 & 1 \end{pmatrix} \qquad (2)$$

Similarly *the properties of the linear* substitution

$$\begin{aligned} x &= x' \cos\theta - y' \sin\theta \\ y &= x' \sin\theta + y' \cos\theta \end{aligned}$$

which describes a rotation of axes of a Cartesian coordinate system, are completely determined by the square 2 by 2 matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

If $m = n$ in (1), A is called a square matrix of order n. If $m = 1$, A is a row matrix, and if $n = 1$,

A is a column matrix. The elements a_{ij} of A, for which $i = j$, are the principal diagonal elements. A diagonal matrix is a matrix such that $a_{ij} = 0$ if $i \neq j$, and a scalar matrix is a square diagonal matrix with equal diagonal elements. An identity matrix is a scalar matrix in which the common diagonal element is the number 1. An n by n identity matrix is denoted by I_n.

Matrices can be regarded as generalized numbers, and their utility in applications depends on the possibility of combining them in certain definite ways. The matrix operations of addition, subtraction, and multiplication are defined in terms of these same operations for the elements, and they satisfy some, but not all, of the rules of ordinary algebra. In discussing these matrix operations it is assumed that the elements of the matrices are numbers.

Two matrices $A = (a_{ij})$ and $B = (b_{ij})$ are equal if they have the same size m by n and $a_{ij} = b_{ij}$ for all i, j. Matrices $A = (a_{ij})$ and $B = (b_{ij})$ of the same size m by n are added by adding correspondingly placed elements; that is, $A + B = C = (c_{ij})$ is an m by n matrix where $c_{ij} = a_{ij} + b_{ij}$ for all i, j. It follows that matrix addition is associative and commutative, that is, $(A + B) + C = A + (B + C)$ and $A + B = B + A$, as is the case for numbers. The m by n matrix with 0 in every position is denoted by 0, and is called a null matrix. Then $A + 0 = 0 + A = A$. The matrix $-A = (-a_{ij})$ is the negative of the matrix $A = (a_{ij})$, and $A + (-A) = 0$. Subtraction of m by n matrices is defined by $B - A = B + (-A) = (b_{ij} - a_{ij})$.

A matrix $B = (b_{ij})$ is conformable with respect to a matrix $A = (a_{ij})$ if B has size n by q and A has size m by n; that is, B has the same number of rows as A has columns. The matrix product AB is defined only when B in conformable with respect to A. The product $C = AB$ is an m by q matrix and the element in the i, j position of C is obtained by multiplying the n elements in the ith row of A into the n elements in the jth column of B, term by term, and adding these products. Thus if

$$A = \begin{pmatrix} 2 & 0 & -1 \\ 4 & 1 & \frac{1}{2} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 5 \\ 1 \\ -2 \end{pmatrix}$$

then

$$AB = \begin{pmatrix} 12 \\ 20 \end{pmatrix}$$

If A and B are square matrices of the same size, then both AB and BA are defined. Matrix multiplication is associative. If A is m by n, B is n by q and C is q by r, then $(AB)C$ and $A(BC)$ are equal m by r matrices. Also, when the matrices A, B, and C have the proper sizes for the operations to be defined, $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$. If A is m by n, then for identity matrices of the proper sizes, $AI_n = I_mA = A$. Unlike the case for numbers, it may happen for matrices that $AB \neq BA$ and $AB = 0$ with $A \neq 0$ and $B \neq 0$.

The product of a matrix A and a number a is called a scalar product and is obtained by multiplying every element a_{ij} of A by a. Thus,

$$\tfrac{1}{2}\begin{pmatrix} 4 & -3 & \frac{1}{3} \\ 0 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 2 & -\frac{3}{2} & \frac{1}{6} \\ 0 & 1 & \frac{1}{2} \end{pmatrix}$$

The transpose of an m by n matrix A is the n by m matrix B which has as its ith row the ith column of A and as its jth column the jth row of A for all i, j. If the transpose of a matrix A is denoted by A', and B is conformable with respect to A, then $(AB)' = B'A'$. A matrix A is symmetric if $A = A'$. A symmetric matrix is necessarily square.

A square n by n matrix is nonsingular if the determinant (*see* DETERMINANT) of A is not zero. Otherwise A is singular. A nonsingular matrix A has a unique inverse, that is, a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$. The inverse of a nonsingular matrix is easily described but is difficult to compute for matrices of large size. Many important applications require the calculation of the inverse, and numerical methods are used to approximate the elements of the inverse.

Two m by n matrices A and B are equivalent if there exist nonsingular matrices P and Q such that $B = PAQ$. For an application of equivalence of matrices *see* LINEAR SYSTEMS OF EQUATIONS. The matrices A and B are equivalent if, and only if, B can be obtained from A by a sequence of elementary transformations which consists of the following operations: interchanging two rows (or columns) of A; multiplying the elements of a row (or column) of A by a fixed number and adding to the corresponding elements of another row (or column) of A; multiplying the elements of a row (or column) of A by a nonzero number. A matrix A can be carried into a matrix with r ones on the principal diagonal and zeros elsewhere by a sequence of elementary transformations. The number r is called the rank of A, and the process of reducing A to this diagonal form is called the reduction of A to canonical form. The rank of A can be defined intrinsically as the largest order r of a nonvanishing minor of A. For example, the matrix (2) has rank 3 since the 3-rowed minor

$$\begin{vmatrix} -3 & 2 & -6 \\ 7 & -1 & 3 \\ 1 & 1 & -5 \end{vmatrix} = 22 \neq 0$$

and the canonical form of this matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Two square n by n matrices A and B are similar if there exists a nonsingular matrix P such that $B = PAP^{-1}$. A linear transformation, or substitution,

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{aligned} \tag{3}$$

can be written $y = Ax$, where y and x are n by 1 column matrices and $A = (a_{ij})$ is the matrix of the

transformation. If new variables $z = Py$ and $w = Px$, where P is nonsingular, are substituted, $P^{-1}z = AP^{-1}w$ or $z = PAP^{-1}w$ is obtained. Thus the given linear transformation in terms of the new variables has a matrix which is similar to A. The theory of a single linear transformation is given by finding canonical forms for the matrix of the transformation under similarity.

If $A = (a_{ij})$ is an n by n square matrix and x is a variable, the matrix

$$Ix - A = \begin{pmatrix} x - a_{11} & a_{12} & & a_{1n} \\ a_{21} & x - a_{22} & & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & & x - a_{nn} \end{pmatrix}$$

is called the characteristic matrix of A. The determinant $|Ix - A|$ of $Ix - A$ is a polynomial in x of degree n, and the equation $|Ix - A| = 0$ is called the characteristic equation of A. The roots of the characteristic equation are the characteristic values, or *eigenvalues*, of A. Many applications in mathematics and physics require information about the characteristic values of a matrix.

A quadratic form in n variables $x_1, x_2, \ldots, x_n$ is a polynomial of degree 2 which can be written as the matrix product xAx', where $x = (x_1, x_2, \ldots, x_n)$ [illegible] metric mat [illegible] made, the [illegible] singular, the new matrix PAP' is said to be congruent to A. The simplification of a quadratic form [illegible] congruence transformations, the particular form depending on the number system (that is, rational, real, or complex numbers) which contains the elements of A and the elements of the transforming matrix P. When A has been reduced to diagonal form, the corresponding quadratic form is a sum of squares of the new variables $y_1, y_2, \ldots, y_n$.

A square matrix A with complex number ele- [illegible] hermitian matrix to diagonal form by a matrix transformation of the form PAP^{*} is called a conjunctive reduction.

A nonsingular matrix P with real number elements is called orthogonal if $P' = P^{-1}$. The congruence transformation which replaces a square matrix A with real number elements by PAP', where P is an orthogonal matrix, is called an orthogonal transformation. If A is a real symmetric matrix, then A can be reduced by an orthogonal transformation to a diagonal matrix which has the n characteristic values of A on the diagonal. The characteristic values of a hermitian matrix and a real symmetric matrix are real numbers.

A nonsingular matrix P with complex number elements is called unitary if $P^{*} = P^{-1}$. The similarity transformation PAP^{-1} where P is unitary is a unitary transformation. The diagonal form obtained for a hermitian matrix A by a unitary transformation has the characteristic values of A on the diagonal. [R.A.B.]

Bibliography: R. A. Beaumont and R. W. Ball, *Introduction to Modern Algebra and Matrix Theory*, 1954; R. R. Stoll, *Linear Algebra and Matrix Theory*, 1952; H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices*, 1932.

Matter (physics)

The substance comprising bodies perceptible to the senses. The distinguishing properties of matter are gravitation and inertia (see GRAVITATION; INERTIA). Any entity exhibiting these properties when at rest is matter. Although electromagnetic radiation also possesses these properties to some extent, this radiation always moves with the speed of light. All material bodies have mass, which is a measure of inertia; every material body near the earth's surface has weight, which is a measure of the earth's gravitational attraction for the body (see MASS; WEIGHT).

According to Einstein's theory of relativity, matter and energy are equivalent quantities; associated with matter of mass m, there is energy E given by the expression $E = mc^2$, where c is the speed of light. See INERTIA OF ENERGY. [D.W.I.]

Matterhorn

A high isolated peak (14,780 ft) with unusually steep sides in southern Switzerland near Zermatt; also called Mont Cervin and Monte Silvio. The name, sometimes shortened to horn, is used as a type designation for other peaks of the same origin, namely sapping of the headwalls of opposite cirques leaving a steep residual mountain. The con-



The Matterhorn, a peak in the Pennine Alps on the Swiss-Italian border. (From Library of Congress Collections)

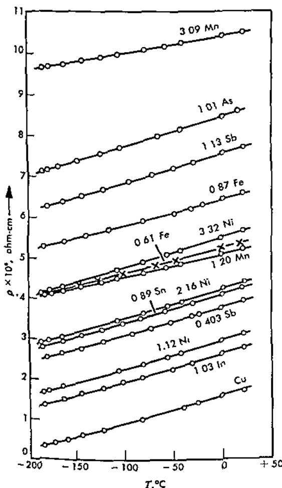
trast of such peaks with adjacent unglaciated mountains is evidence of the nature of glacial erosion. *See* GLACIATED TERRANE. [F.T.T.]

Matthiessen's rule

An empirical rule which states that the total resistivity of a crystalline metallic specimen is the sum of the resistivity due to thermal agitation of the metal ions of the lattice and the resistivity due to the presence of imperfections in the crystal. Matthiessen's rule provides a basis for the understanding of the behavior of the resistivity of metals and alloys at low temperatures.

The resistivity of a metal results from the scattering of conduction electrons (*see* RESISTIVITY, ELECTRICAL). Lattice vibrations scatter electrons because the vibrations distort the crystal. Imperfections such as impurity atoms, interstitials, dislocations, and grain boundaries scatter conduction electrons because in their immediate vicinity the electrostatic potential differs from that of the perfect crystal. If lattice scattering and imperfection scattering are independent processes and are isotropic, it can be shown that

$$\rho(T) = \rho_L(T) + \rho_I \qquad (1)$$



Resistivity of copper and various copper alloys. (*After* J. O. Linde)

where $\rho(T)$ is the resistivity at temperature T, $\rho_L(T)$ is the resistivity due to lattice scattering (the ideal resistivity, which is temperature-dependent), and ρ_I is the so-called residual resistivity due to imperfections, which is presumably independent of temperature. An equivalent statement is

$$\frac{\partial\rho(T)}{\partial T} = \frac{\partial\rho_L(T)}{\partial T} \qquad (2)$$

As exemplified by the illustration, Matthiessen's rule is generally obeyed, although some deviations do occur. Such deviations arise for the following reasons: (1) the introduction of impurities generally alters the elastic properties, influences the lattice vibration spectrum, and thereby changes $\rho_L(T)$; (2) some imperfections, such as dislocations, do not scatter isotropically; (3) thermal expansion lowers the Fermi energy (since impurity scattering depends on the velocity of the electrons at the Fermi energy, the resistivity due to impurities depends also on the temperature through the thermal expansion); and (4) the impurity ions also participate in the thermal vibrations of the crystal lattice (however, since the scattering potential of an impurity differs from that of a solvent ion, lattice scattering from the former also differs from that due to a solvent ion). [F.J.B.]

Bibliography: See RESISTIVITY, ELECTRICAL.

Maxwell

A centimeter-gram-second (cgs) electromagnetic unit of magnetic flux. From the defining equation for magnetic induction,

$$B = F/Il \sin\theta$$

where F is the force on a current-carrying element of length l carrying a current I and making an angle θ with the direction of B. The cgs unit of B is the dyne/abampere-centimeter and is called the gauss. *See* GAUSS; INDUCTION, MAGNETIC; MAGNETIC FLUX.

In the electromagnetic system of units, the lines of flux representing the magnetic induction are chosen so that the number of lines per square centimeter of a surface perpendicular to B is equal to the value of B. The maxwell is one line of induction as thus defined. A gauss is thus a maxwell per square centimeter. *See* ELECTRICAL UNITS. [K.V.M.]

Maxwell's equations

Four differential equations proposed by James Clerk Maxwell in 1864 as the basis of the theory of electromagnetic waves. They may be written, in vector notation,

$$\nabla \cdot \mathbf{D} = \rho \qquad (1)$$

$$\nabla \cdot \mathbf{B} = 0 \qquad (2)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad (3)$$

$$\nabla \times \mathbf{H} = \mathbf{i} + \frac{\partial \mathbf{D}}{} \qquad (4)$$

where **D** is the electric di ... **B** the netic flux density, **E** the ... l stre

intensity, H the magnetic field strength or intensity, ρ the charge density, and i the current density.

The first equation states that electric flux lines, if they end at all, will do so on electric charges. The second states that magnetic flux lines never terminate. The third is a form of Faraday's law of induction, which states that the rate of change of the magnetic flux threading a circuit equals the electromotive force or line integral of E around the circuit so that

$$-\frac{\partial}{\partial t}\int_S \mathbf{B}\cdot\mathbf{n}\,dS = \oint \mathbf{E}\cdot d\mathbf{s} = \int_S \nabla\times\mathbf{E}\cdot\mathbf{n}\,dS \quad (5)$$

where n is a unit vector normal to the surface S. The third integral comes from the second by Stokes' theorem (see STOKES' THEOREM). If this holds for any surface S, then the integrands of the surface integrals are equal and Eq (3) follows. The fourth integral is based partially on A. M. Ampère's experiments on steady currents which show that the line integral of the magnetic intensity H (or B/μ, where μ is the permeability) around a closed curve equals the current encircled, which is found by integration of the normal current density over the enclosed area. Thus

$$\int_S \mathbf{i}\cdot\mathbf{n}\,dS = \oint \mathbf{H}\cdot d\mathbf{s} = \int_S \nabla\times\mathbf{H}\cdot\mathbf{n}\,dS \quad (6)$$

Stokes' theorem transforms the second integral into the third. The surface S is arbitrary so the integrands of the surface integrals are equal, and this gives Eq. (4) without the term $\partial D/\partial t$. Maxwell realized that for fluctuating currents, where charges *may accumulate*, the $\partial D/\partial t$ term was needed to satisfy the equation of continuity which requires that, if charge is to be *conserved*, its rate of increase in any region must equal the flow of current into that region so that

$$\nabla\cdot\mathbf{i} + \frac{\partial\rho}{\partial t} = 0 \quad (7)$$

When the divergence of Eq. (4) is taken and D is eliminated by Eq. (1), Eq. (4) results. The quantity $\partial D/\partial t$ is called the displacement current. *See* DISPLACEMENT CURRENT; EQUATION OF CONTINUITY.

Rectangular coordinates. When the components of Eqs. (1) to (4) are written in rectangular coordinates the result is

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho \quad (8)$$

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0 \quad (9)$$

$$\frac{\partial E_{z,x,y}}{\partial u_{y,z,x}} - \frac{\partial E_{y,z,x}}{\partial u_{z,x,y}} = -\frac{\partial B_{x,y,z}}{\partial t} \quad (10)$$

$$\frac{\partial H_{z,x,y}}{\partial u_{y,z,x}} - \frac{\partial H_{y,z,x}}{\partial u_{z,x,y}} = i_{x,y,z} + \frac{\partial D_{x,y,z}}{\partial t} \quad (11)$$

In Eqs. (10) and (11), the *first, second, or third* subscript must be used throughout, depending on the component desired, and u_x, u_y, and u_z are identical with x, y, z.

Ratio of units. In Maxwell's time, E and B were measured in a different system of units from E and D, the ratio of the units of charge being c. In a vacuum, where i is zero, this introduces a factor $1/c$ on the right sides of Eqs. (3) and (4), so that when Eq (4) is used to eliminate the curl of B from the curl of Eq. (3), there results

$$\nabla\times\nabla\times\mathbf{E} = \nabla^2\mathbf{E} = \frac{1}{c}\frac{\partial^2\mathbf{E}}{\partial t^2} \quad (12)$$

The first and second terms are equal because $\nabla\cdot E$ is zero in a vacuum. This is the differential equation for a wave of velocity c. Thus, the experimental verification of Maxwell's prediction that the velocity of light equals the ratio of the electromagnetic unit to the electrostatic unit of charge proves the electromagnetic nature of light. *See* ELECTRICAL UNITS; ELECTROMAGNETIC RADIATION; LIGHT; WAVE EQUATION.

Conducting media. If γ is the conductivity of a substance, Ohm's law states that the current density i in Eq. (4) may be replaced by γE. If this is done and H and D are replaced by B/μ and ϵE, then the use of Eq. (3) to eliminate the curl of E from the curl of Eq. (4) gives

$$\nabla^2\mathbf{B} = \mu\gamma\frac{\partial\mathbf{B}}{\partial t} + \mu\epsilon\frac{\partial^2\mathbf{B}}{\partial t^2} \quad (13)$$

which is the differential equation for a damped wave. This equation also holds if E replaces B. When γ is much larger than ϵ, the second term on the right may be dropped so that Eq. (13) becomes the differential equation for eddy currents used in skin-effect and induction-heating calculations.

Index of refraction. When γ is zero, the solutions of Eq. (13) represent undamped waves of velocity $(\mu\epsilon)^{-1/2}$. In optics, the index of refraction of a medium is defined as the ratio of the velocity of light in a vacuum to that in the medium. Thus, by Maxwell's theory, the index of refraction n may be expressed in terms of the permeability and the dielectric constant of the medium by the relation

$$n = \frac{c}{v} = \left(\frac{\mu\epsilon}{\mu_0\epsilon_0}\right)^{1/2} \quad (14)$$

where μ_0 and ϵ_0 are the vacuum values of the permeability and the dielectric constant, and v is the velocity of light in the medium. For everything except ferromagnetic materials, μ is approximately equal to μ_0, so these symbols cancel. The dielectric constant ϵ is a function of frequency, so static values should be used with caution, although they often hold in the radio-frequency range. *See* REFRACTION OF WAVES.

Integral form. In some ways, the integral forms of Eqs. (1) to (4) give a clearer physical concept than the differential forms. With the aid of Eqs (5) and (6), Maxwell's equations may be written

$$\int \mathbf{D} \cdot \mathbf{n}\, dS = \int \rho\, dv \tag{15}$$

$$\int \mathbf{B} \cdot \mathbf{n}\, dS = 0 \tag{16}$$

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot \mathbf{n}\, dS \tag{17}$$

$$\oint \mathbf{H} \cdot d\mathbf{s} = \int_S \left(\mathbf{i} + \frac{\partial \mathbf{D}}{\partial t} \cdot \mathbf{n}\right) dS \tag{18}$$

Lorentz invariance. The form of Maxwell's equations is the same for all observers whose coordinate systems move with a uniform translational velocity relative to each other. The values of the observed fields, as well as ρ, $\mathbf{i}$, μ, ϵ, and γ, may be quite different. This Lorentz transformation can sometimes be used to remove one of the fields entirely, thus greatly simplifying the calculation. *See* CALCULUS OF VECTORS; LORENTZ TRANSFORMATIONS. [W.R.SM.]

Bibliography: J. C. Maxwell, *Treatise on Electricity and Magnetism*, reprint, 1954; J. A. Stratton, *Electromagnetic Theory*, 1941; E. T. Whittaker, *A History of the Theories of the Aether and Electricity*, vol. 1, 1951.

Mayfly

Any member of the insect order Ephemeroptera (Ephemerida). Mayflies are characterized by two pairs of frail, membranous, net-veined wings, the front pair larger than the hind pair, and all triangular in shape; two or three long, filamentous appendages at the end of the abdomen, soft bodies; incomplete metamorphosis; and vestigial mouthparts in the adult. The nymphs are aquatic, usually with paired tracheal gills on the abdomen, and chewing mouthparts. Mayflies are famous for the swarms of adults which sometimes emerge in such great numbers that they block traffic on bridges or streets near water. They usually live as nymphs in clear water for 1–4 years before emerging. The nymphs eat small plants and decaying vegetable matter. The adults do not eat and usually live only a few hours. They are significant as fish food in many places. There are about 1500 known species, 500 in North America. *See* EPHEMEROPTERA. [J.D.B.]

Mealybug

Any of several species of insects of the family Coccidae, order Homoptera. This family also includes the scale insects. Mealybugs derive their name from the fluffy, waxy powder with which they are covered. They, as well as some of the scale insects, produce honeydew, which may damage plants because of the mold which grows on it. Mealybugs also damage plants by feeding on the sap. Only the males are winged. Some species lay eggs, whereas others produce living young.



Citrus mealybug; *Pseudococcus citri*; length to ¼ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Several are of economic importance; the citrus mealybug, *Pseudococcus citri*, for instance, attacks citrus, ornamental, and flowering plants, and is a major greenhouse pest. *See* HOMOPTERA. [J.D.B.]

Mean free path

The average distance traveled between two similar events. The concept of mean free path is met in all fields of science and is classified by the events which take place. The concept is most useful in systems which can be treated statistically, and is most frequently used in the theoretical interpretation of transport phenomena in gases and solids, such as diffusion, viscosity, heat conduction, and electrical conduction. The types of mean free paths which are used most frequently are for elastic collisions of molecules in a gas, of electrons in a crystal, of phonons in a crystal, and of neutrons in a moderator.

An elementary formula for the mean free path for elastic collision of a molecule in a gas can be derived in the following way. In a gas at a pressure P, let n be the average number of molecules [illegible]

rest. In a time t, the moving molecule will "sweep out" a volume $4\pi a^2 ct$, and since there are n molecules per unit volume, there will be $4\pi a^2 ctn$ collisions. The distance traveled is ct; therefore the average distance between collisions is $1/(4\pi a^2 n)$. Other more exact methods of taking averages give values which are closer to those determined experimentally. In hydrogen at normal temperature and pressure the mean free path is 1.7×10^{-5} cm. *See* KINETIC THEORY OF MATTER. [W.D.W.]

Meantes

A suborder of the Caudata or salamanders. There are two genera of these amphibians, *Pseudobranchus* with a single species and *Siren* with two species. Commonly known as mud-eels, these salamanders are found in areas of low elevation from Texas to Florida, Maryland, and Illinois. The mud-eels are neotenic, retaining such larval characters as lidless eyes and external gills. The body form is eel-like, with only anterior limbs being present. *Siren* reaches a length of 36 in., while *Pseudobranchus* is much smaller, less than 9 in. long. Members of both genera are wholly aquatic, and are found beneath rocks in streams, or burrow i the muck and vegetation of bogs and swamps. CAUDATA; NEOTENY. [

Measles

An acute, highly infectious viral disease, with cough, fever, and maculopapular rash. It is of world-wide endemicity. See ANIMAL VIRUS.

The infective particle is 140 mμ, measured by ultrafiltration, but the active core is only 65 mμ as measured by inactivation after electron irradiation. It will infect monkeys easily, chick embryos with [illegible] in tissue cultures the virus may produce [illegible] dophilic [illegible] ED EGG; CULTURE, TISSUE; INCLUSION BODIES (VIRUS).

The virus enters the body via the respiratory system, multiplies there, and circulates in the blood. Prodromal cough, sneezing, conjunctivitis, photophobia, and fever occur, with Koplik's spots appearing in the mouth. A rash appears after 14 days' incubation and persists 5–10 days. Encephalomyelitis is rare.

Laboratory diagnosis, although seldom needed since 95% of cases have the pathognomonic Koplik's spots, is by virus isolation in tissue cultures from acute-phase blood or nasopharyngeal secretions, or by complement-fixing antibody responses. See COMPLEMENT-FIXATION TEST.

Immunizing infections occur in early childhood, during epidemics which recur after 2–3 years' accumulation of susceptible children. Transmission is by coughing or sneezing. By the age of 20 years, over 80% of persons have had the disease. Treatment is symptomatic. Gamma globulin given early can modify or prevent the disease, depending on the chosen dosage. Prevention is advisable in infants 4–12 months old and in children sick with other disease. See GAMMA GLOBULIN. [J.L.M.]

Measure

A reference sample used in comparing lengths, areas, volumes, masses, and the like. The measures employed in scientific work are all based directly on the fundamental *international* units of length, mass, and time—the meter, the *kilogram*, and the *second*; only decimal multiples and submultiples are commonly employed in scientific work. However, prior to the development of the *international* metric system, many special-purpose *systems of measures* had evolved and many still survive today, especially in Great Britain and the United States, where metric units have not come into common use (see METRIC SYSTEM). The multiples and submultiples in these special-purpose systems are usually nondecimal; some are binary, based on successive doubling or halving; others involve subdivision of 60 or 12; still others have strange multiples and submultiples of ancient and sometimes unknown origin. A few common systems are listed here; others can be found in engineering handbooks.

Length measures

12 inches = 1 foot (= 0.3048 meters)
3 feet = 1 yard
5½ yards = 1 rod
40 rods = 1 furlong = 220 yards
8 furlongs = 1 mile = 5280 feet
4 inches = 1 hand
9 inches = 1 span

Surveyor's length measures

7.92 inches = 1 link
100 links = 1 chain
10 chains = 220 yards = 1 furlong
8 furlongs = 1 mile

Nautical length measures

6080.26 feet = 1.15156 statute miles = *1 nautical mile*
3 nautical miles = *1 league*
60 nautical miles = 1 degree at Equator
6 feet = *1 fathom*
120 fathoms = *1 cable*

Measures of area

1 acre = 160 square rods = 4840 square yards
640 acres = 1 square mile
1 circular inch = area of circle 1 inch in diameter = 0.7854 square inches
1 circular mil = area of circle 0.001 inch in diameter

Measures of mass (avoirdupois weight)

16 grams = 1 ounce
16 ounces = 7000 grains = 1 pound (= 453.6 grams)
14 pounds = 1 stone
8 stones = 112 pounds = 1 hundredweight (cwt)
20 cwt = 2240 pounds = 1 long ton
2000 pounds = 1 short ton

Troy weight (for gold and silver), carat weight (for precious stones), and apothecaries' weight (for drugs) are other systems, in all of which the grain is the same as the avoirdupois grain.

Measures of volume

Liquid measure

4 gills = 1 pint
2 pints = 1 quart
4 quarts = 1 U.S. gallon = 231 cubic inches
1 Imperial gallon = 1.20094 U.S. gallons

"Old Liquid Measures" involving 1 barrel = 31½ gallons and apothecaries' measures involving 16 fluid ounces = 1 pint are still in wide use.

Dry measure

2 pints = 1 quart
8 quarts = 1 peck
4 pecks = 1 bushel = 2150.42 cubic inches
105 quarts = 1 barrel
8 Imperial gallons = 1 Imperial bushel

Circular measure

60 seconds = 1 minute
60 minutes = 1 degree
90 degrees = 1 quadrant

360 degrees = 1 revolution
2π radians = 1 revolution

Time

60 seconds = 1 minute
60 minutes = 1 hour
24 hours = 1 day
7 days = 1 week
365 days, 5 hours, 48 min, 45.6 sec = 1 year
1 sidereal day = 86,164 sec

See TIME; WEIGHT. [D.WI.]

Bibliography: Units and Systems of Weights and Measures, Natl. Bur. Standards Circular 570, 1956.

Mechanical advantage

Ratio of the force exerted by a machine (the output) to the force exerted on the machine, usually by an operator (the input). The term is useful in discussing a simple machine, where it becomes a figure of merit. It is not particularly useful, however, when applied to more complicated machines where other considerations become more important than a simple ratio of forces. *See* EFFICIENCY; SIMPLE MACHINE. [R.M.PH.]

Mechanical engineering

One of several recognized fields of engineering. This distinction begins in college where, after one year of common engineering studies, a person identifies himself as a student of mechanical engineering. After college, mechanical engineers may enter many industries. Common areas of industrial employment for mechanical engineers are private electric power and machinery manufacturing.

To grasp the meaning of mechanical engineering, it is desirable to take a close look at what engineering really is. The Engineers' Council for Professional Development has defined engineering as the profession in which a knowledge of the mathematical and physical sciences gained by study, experience, and practice is applied with judgment to develop ways to utilize, economically, the materials and forces of nature for the progressive well-being of mankind. Here is a picture of a profession in which study in mathematics and science is blended with experience and judgment toward the end of making useful things.

Formal training of a mechanical engineer includes mastery of mathematics through the level of differential equations. His training in physical science embraces chemistry, physics, mechanics of materials, fluid mechanics, thermodynamics, statics, and dynamics. Enhancing these subjects are courses in the humanities: literature, economics, philosophy, and history.

Experience, for the mechanical engineer, is not gained solely by the passage of time. For experience to be added, the time spent in engineering must be meaningful. Projects must become more difficult and the consequences of error greater.

Judgment is the hallmark of any competent engineer, in whatever recognized branch of engineering he specializes. The exercise of judgment requires the ability to assess and decide between alternate courses of action. In the application of judgment, there is for the engineer constant testing of his decisions against his knowledge of the laws of nature, his sense of right and wrong, and economics.

A further essential of an engineer is that his efforts be devoted to matters which improve the well-being of mankind. This purpose is kindred to that of all learned professions and is closely allied to the professions of medicine and law.

In relation to other professions, the mechanical engineer differs in one large measure. Whereas most doctors and many lawyers are self-employed, almost all engineers are employed by corporations, colleges, or government. This distinction exists for several reasons. Generally, the doctor or lawyer is far more concerned with people and their actions than is the engineer. Further, an engineer dealing with machines requires larger sums of money to finance the equipment and facilities he uses than does the doctor or lawyer.

From the foregoing, a young person thinking of becoming a mechanical engineer should recognize that there is a long period of preparation ahead. Starting in high school, he should pursue a full rigorous academic program. Then, careful self-examination of his results in high school are in order. If there is demonstrated high interest and success in mathematics, English, and science, there is room to consider the matter further. Finally, there must be evidence of willingness to work and patience to gain the experience and judgment required in any learned profession, including mechanical engineering. *See* ENGINEERING; MACHINE DESIGN; MACHINERY; MASS PRODUCTION; TECHNOLOGY. [R.S.SH.]

Mechanical rectifier

A device which uses a synchronously operated mechanical switch to convert a single-phase or polyphase alternating voltage to a direct voltage. Single-phase mechanical rectifiers are made for small current output and are normally called vibrators (*see* VIBRATOR). For large values of power where low voltages (less than 600 volts) are desired, the polyphase mechanical rectifier is used. This low-voltage device has higher efficiency than electronic rectifiers, which have appreciable voltage drop across the arc. *See* MERCURY-VAPOR RECTIFIER.

Figure 1 shows the elementary circuit of the mechanical rectifier used on a three-phase system. The transformer connections are similar to those of other rectifiers. Nonlinear reactors L_1, L_2 and L_3 are used in the output leads of the transformer

between the vertical portions of the loop. The function of these reactors is twofold: (1) they pro-

vide good commutation at the contacts C_1, C_2, C_3, and (2) they serve to limit short-circuit current when two of the segments are in contact with the rotating member of the synchronous switch.

The rotating element of the synchronous switch is driven by a synchronous motor which operates from the same ac source as the rectifier transformers The output voltages of the transformers have a time sequence of e_{n1}, e_{n2}, and e_{n3}. Contacts between the rotating element of the switch and the stationary outer element are closed in the order of C_1, C_2 and C_3. When contact is made between C_1 and the rotating element, current I_1 is supplied from e_{n1} to the load. At the instant t_1 (see Fig. 3) the voltage e_{n2} becomes larger than e_{n1}, the rotating element makes contact to C_2, and the current I_2 increases while I_1 decreases. During the interval t_1 to t_2 that C_2 and C_1 are short circuited by the ... I_c will flow as shown in

cle by L_1, the inductance of L_1 becomes large. At this instant t_2, the rotating element of the switch breaks contact with C_1. Thus the contact is broken when the current from C_1 to the rotating element

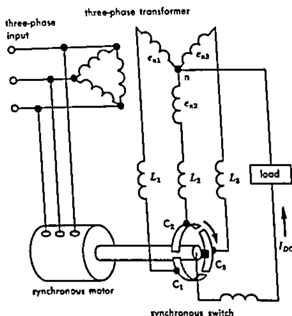


Fig 1. Mechanical rectifier circuit.

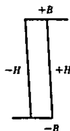


Fig. 2. Hysteresis loop of inductor cores.

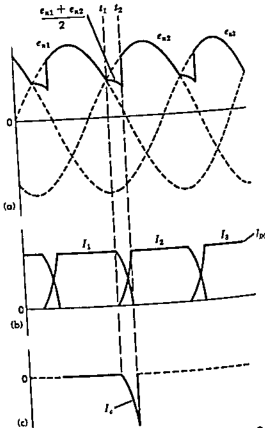


Fig. 3. Voltage and current wave shapes. (a) Output voltage. (b) Transformer secondary currents. (c) Circulating current I_c.

is essentially zero. This provides for good commutation.

The output voltage wave of such a rectifier, the circulating current during the interval of commutation, and the currents of each of the transformer secondaries are shown in Fig. 3. The efficiency of these devices may be as high as 97%. See RECTIFIER. [A.C.CO.]

Mechanical vibration

A motion, usually unintentional and often undesirable, of parts of machines and structures. Vibrations occur frequently in rotating as well as in stationary machines and structures. Although these motions and their causes are sometimes highly complicated, a large number of cases can be understood from the study of one of the prototypes (Fig. 1). These prototypes are systems with a single degree of freedom of the linear type as well as of the torsional or rotational type (Fig. 1). They contain a mass or weight W (moment of inertia I in the torsional case), a spring k driving the mass to a neutral position, a damper or dashpot c opposing the velocity of motion, and an alternating exciting force (or torque) $P = P_{max} \sin 2\pi ft$. In the absence of excitation P and damping c, the system can execute free, undamped vibrations of natural

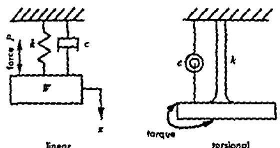


Fig. 1. Vibrating systems with one degree of freedom.

frequency

$$f_n = \frac{1}{2\pi}\sqrt{gk/W}$$

where f_n is measured in cycles per second; k is the spring constant, measured in pounds per inch deflection; W is the weight in pounds; and $g = 386$ in./sec².

If a small damping c is added to these (nonexcited) systems, the natural frequency remains substantially the same; the motion gradually decays However when damping c becomes large, the system ceases to be vibratory, and the motion is one of creeping back to the equilibrium position, like a door with a doorstop. This value of the damping is called the critical damping c_{crit},

$$c_{crit} = 2\sqrt{kW/g}$$

For the usual case where damping c is much smaller than this critical value it is customary to express it as a ratio to the critical damping c/c_{crit}.

Resonance. When the exciting force (or torque) $P_{max} \sin 2\pi ft$ is added to the system, it is excited to a forced vibration at the forced frequency f, which can be chosen freely. If this forced frequency is varied slowly, the resulting amplitude of motion varies, and when this motion ($x = x_{max} \sin 2\pi ft$), divided by exciting force, $x_{max}/(P_{max}/k)$ is plotted against frequency ratio f/f_n for various values of damping c/c_{crit}, Fig. 2, which is a resonance diagram, results. In general

$$\frac{x_{max}}{P_{max}/k} = \frac{1}{\sqrt{(1 - f^2/f^2_n)^2 + \left(2\frac{f}{f_n}\frac{c}{c_{crit}}\right)^2}}$$

When the forced and natural frequencies are close to each other, the amplitude of motion becomes very large (resonance). The most common and important engineering problems in vibration are (1) to recognize what the disturbing or exciting force P is in magnitude and frequency, (2) to calculate or estimate the natural frequency of the system, and (3) to arrange the design so that these two frequencies are sufficiently far apart to avoid resonance.

The most serious effect of vibration on machinery is that the alternating stress, if high enough, can break the machine by fatigue; less serious effects are increased wear of parts, malfunctioning of apparatus, and the spreading of the vibration through foundations and buildings to locations where vibration is intolerable either for human comfort or for the operation of sensitive measuring equipment.

When the system is more complicated than the prototypes of Fig. 1 and has more than one mass or flywheel there are more degrees of freedom and the resulting resonance diagram has as many natural frequencies and resonant peaks as there are degrees of freedom. However, in most mechanical systems only the first few lowest frequencies of these resonances assume practical importance.

The principal causes of forced vibration in practice are (1) torsional vibrations in shaft systems containing piston engines or propellers, (2) longitudinal vibrations in ship shafts or in liquid pipelines, (3) lateral vibrations of shafts and rotors caused by the centrifugal force of unbalance, and (4) vibrations of a host of other configurations excited by a variety of causes and idealized as beams, membranes, plates, or rings.

Torsional vibration. An engine-driven system (usually diesel or internal combustion powering a generator, a ship's propeller or other load) has many degrees of freedom and hence many natural frequencies of which the lowest two or three are of practical importance. The installation may, if excited by alternating components in its torque, execute torsional vibrations between its various parts of a magnitude of a fraction of a degree. This relatively small alternating motion is superposed on and independent of the continuous rotation of the engine shaft. Whereas the continuous rotation causes no extreme stress in the shaft, the small (quarter degree) angles of vibration wind up the

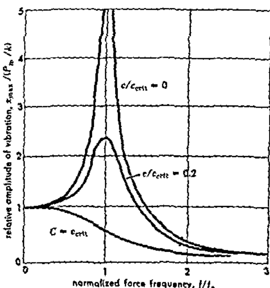


Fig. 2. Amplitude of vibration varies with forced frequency and degree of damping.

shaft and relax it causing alternating stresses that on many occasions have caused failure.

The torque on a crankshaft caused by a single cylinder and piston having one explosion for each one or two revolutions has a highly irregular time history containing many harmonics. The torques developed by the various cylinders of a multicylinder engine combine in accordance with their times of occurrences. Thus the exciting torque contains components of the firing frequency (one cycle per one or two revolutions) and multiples of that firing frequency, which are of practical importance up to the 16th multiple or harmonic. Because these 16 exciting or forced frequencies are proportional to the engine speed and the two or three natural frequencies are independent of it, and each exciting frequency can resonate with each natural one, there are as many as 48 resonances or critical speeds to be considered. Many of these will lie outside the habitual running range of the engine.

The designer may shift the severity of the resonances, by using flexible couplings or extra flywheels in the engine, or by changing the arrangement of the cylinders, the V-angle of the engine, and the firing order. These suffice to make an installation satisfactory for any one running speed, but it is usually not possible to avoid all dangerous resonances in a wide range of running speeds. The designer must then resort to dampers to keep the amplitude of motion and the stress down. Various types of dampers exist; the most familiar is the pulley in the front of an automobile engine which drives the fan belt. This pulley is usually a small flywheel coupled to the engine shaft through a rubber insert The assembly serves simultaneously as a torsional spring and as a dashpot damper. The flywheel is so tuned that it holds the principal torsional critical speed of the engine below dangerous torsional stress levels. Fatigue failures of crankshafts and of other shafting due to torsional vibration were common in the past, but such failures are avoidable by proper design and are now rare.

By the principle of action equals reaction, the gas torque acting on the piston-crankshaft assembly is equal and opposite to the gas torque acting on the engine frame. Hence, when the engine frame is rigidly attached to a foundation, that foundation experiences an alternating torque. To protect the foundation and to prevent the vibration from spreading through the structure, the engine is often mounted on metal springs or on rubber bushings. This is now universal practice with automobile and aircraft piston engines.

Longitudinal vibration. A large ship's shafting excited by the propeller blades may come to resonance. The shafting and propeller must be designed to allow for such resonance. This is one cause of vibration in passenger liners. It occurs at a frequency of the shaft revolutions multiplied by the number of propeller blades.

The air or gas column in the suction or discharge lines of internal combustion engines acted upon by the piston motions can come to resonance at certain speeds. This principle has been used to increase the power of the engine by designing those lines so that during the intake more air than usual enters the cylinder, and during exhaust more air than usual goes out, thereby decreasing the back pressure. See INTERNAL COMBUSTION ENGINE; RAMJET.

In internal combustion engines the gas pressure exerted on the piston and thence on a crank throw tends to lengthen the crankshaft. Because this effect is periodic, longitudinal vibrations in internal combustion engine shafts have been observed; however, they are of less importance than the torsional ones and become serious only if a longitudinal and a torsional natural frequency fall close together causing coupled oscillations.

Lateral vibrations. The most important of all machinery vibrations are the lateral or bending vibrations of shafts and other rotors caused by the centrifugal force of unbalance. The unbalance consists of a very small deviation of the center of gravity of the rotor from the geometric axis connecting the bearing centers, or of a small angular deviation between that bearing center line and a principal axis of inertia of the rotor. Thus the excitation always has the frequency of the shaft rotation and when that coincides with one of the natural bending frequencies of the rotor there is resonance or a critical speed where the bending stresses in the shaft can be a hundred times higher than those caused by the centrifugal force of the unbalance directly.

Slow speed machines are always designed to run well below the lowest or first critical speed, but for high speed ones this often is not possible. Steam turbines and electric generators of large power stations have for many years been designed to run between the first and second critical speeds; the newest and largest units are between their second and third criticals. The flexibility of the supporting (nonrotating) bearings is an important factor in the calculation of critical speeds or natural frequencies of rotors, because the flexible bearing decreases the natural frequencies by some ten per cent on the average below those that would be present if rigid bearings were used. For rotors of a high diameter-length ratio, such as in steam turbines with large diameter discs mounted on a comparatively thin shaft, the effect of rotating inertia, sometimes called the gyroscopic effect of the discs, is an important consideration.

Besides the classical unbalance critical speed, which always has a vibration frequency equal to that of the rpm, a number of secondary critical speeds have been observed and explained in which the frequency of vibration is a multiple of, usually twice, the rpm. The practical importance of these is secondary with respect to the ordinary critical speed.

Beam vibrations. Whenever a part of a structure or the entire structure itself has a natural frequency resonating with an alternating excitation nearby (or sometimes far removed) severe vibra-

tions may result. A typical example is an unbalanced piston machine, such as an air compressor, which frequently causes objectionable vibration in some locations in the building where it is installed. Another example is the vibration of an entire ship in the mode of a free-free beam, excited by the propeller blade frequency.

With the advent of jet engines, the effect of high intensity airborne noise on the very light structures of airplanes and missiles has become important. It is characteristic of jet noise, or indeed of all cases of turbulent flow, that the excitation is distributed continuously over a wide band of frequencies. This type of excitation is random and is playing an increasingly important role in the design of aircraft and missiles.

Membrane vibration. A tightly stretched skin, which has negligible bending stiffness like a drumhead, is a membrane. (The diaphragm in a telephone receiver possesses considerable bending stiffness, and is not stretched; hence technically it is a plate, although sometimes it is also called a membrane.) The theory of vibration of a membrane of circular shape has been known for a century and is one of the more beautiful illustrations of the mathematics of Bessel functions. The lowest frequency of vibration corresponds to a shape where the entire membrane bulges in and out, without nodal lines, the periphery remaining fixed. The higher modes of motion possess nodal lines, which are concentric circles or angularly equidistant diameters. The frequency formula for the circular membrane vibration is

$$f = a\sqrt{Tg/\gamma r^2}$$

in cycles per second, where T is the tension in the membrane in pounds per inch, $g = 386$ in./sec^2, γ is the unit weight of the membrane in pounds per square inch, r is the radius in inches, and a is a numerical factor, having the values shown in the table.

Numerical factor for membrane frequency

Number of nodal circles	Number of nodal diameters			
	0	1	2	3
0	0.383	0.610	0.819	1.02
1	0.880	1.12	1.34	1.55
2	1.38	1.62	1.85	2.08

Self-excited vibrations. In the cases discussed so far there existed an alternating exciting force (or torque) which would continue to exist by itself even after the vibratory motion was prevented or stopped. These are called forced vibrations. There is another class of motions, called self-excited ones, whereby the exciting force is generated by the vibrating motion itself and hence disappears with that motion.

A familiar mechanical example of self-excited vibration is the piston of an engine. The back and forth motion (vibration) is maintained by an alternating gas or steam pressure steered by the valve mechanism. The initial source of energy is without alternating properties (the gasoline supply or the boiler steam) but is made alternating by the valves, which are operated by the vibratory piston itself.

A familiar electrical example is the self-oscillating electronic circuit where the steady source of energy is the B-battery, and the grid in the vacuum tube plays the role of the valve.

The oldest practical examples of useful self-excited vibrations are musical instruments, particularly the violin and the clarinet. The violin operates on the peculiar behavior of the friction between the rosined bow and the string, which has aptly been described as stick-slip friction. Vibrations of this type have appeared repeatedly and still are often met in machinery with insufficient lubrication; the most vexing and difficult case being that of the chattering machine tool cutter which leaves a wavy cut instead of a smooth one.

The clarinet operates on the passage of air (from the mouth to the instrument) through a narrow leakage opening whose width varies periodically with the vibration of the reed. Serious vibrations of this character occur in steam, gas, or hydraulic turbines, heat exchangers, and other apparatus in which a fluid or gas passes through narrow passages. Other self-excited vibrations are the pulsating flow sometimes observed in fans and blowers and the shimmying motion of wheels which has been a serious problem in automobiles and in the landing gear of aircraft. Self-excited vibrations appear in autopilots of aircraft and missiles, and in general are apt to occur in servomechanisms with high gain.

A class of dangerous vibrations is the various phenomena of flutter, whereby an elastic system becomes self-excited in a stream of air, gas, or fluid of sufficient speed. This aviation problem first arose in airplane wings, but as speeds were increased other parts of airplanes and other machinery, such as turbine blades, displayed flutter. The theory of flutter has grown into a subject by itself (*see* AEROELASTICITY).

Serious vibrations have occurred in rocket engines and other types of combustion chambers, whereby the combustion becomes unstable and the burning gas mass enters a state of self-excited vibration. *These phenomena have been known for a* century and still are not completely understood. *See* VIBRATION. [J.P.D.H.]

Bibliography: J. P. Den Hartog, *Mechanical Vibrations*, 4th ed., 1956; S. Timoshenko and D. H. Young, *Vibration Problems in Engineering*, 3d ed., 1955.

Mechanics

In its original sense, mechanics refers to the study of the behavior of systems under the action of forces. Statics deals with cases where the forces either produce no motion or the motion is not of interest. Dynamics deals properly with motions under forces. Mechanics is subdivided according t the types of systems and phenomena involved.

An important distinction is based on the size of the system. Those systems that are large enough can be adequately described by the Newtonian laws of classical mechanics. In this category, for example, are celestial mechanics, the study of the motions of planets, stars, and other heavenly bodies, and fluid mechanics, which treats liquids and gases on a macroscopic scale. Fluid mechanics is a part of a larger field called continuum mechanics or (by some physicists) classical field theory, involving any essentially continuous distribution of matter, whether rigid, elastic, plastic, or fluid. On the other hand, the behavior of microscopic systems such as molecules, atoms, and nuclei can be interpreted only by the concepts and mathematical methods of quantum mechanics.

From its inception, quantum mechanics had two apparently different mathematical forms—the wave mechanics of E. Schrodinger, which emphasizes the spatial probability distributions in the quantum states, and the matrix mechanics of W. Heisenberg, which emphasizes the transitions between states. These are now known to be equivalent

Mechanics may also be classified as nonrelativistic or relativistic mechanics, the latter [illegible]
system [illegible]
velocit [illegible]
classical and quantum mechanics.

Finally, statistical mechanics uses the methods of statistics for both classical and quantum systems containing very large numbers of similar subsystems to obtain their large-scale properties. *See* CELESTIAL MECHANICS; DYNAMICS; FIELD THEORY, CLASSICAL, FLUID MECHANICS; MECHANICS, CLASSICAL; QUANTUM MECHANICS; RELATIVISTIC MECHANICS; STATICS; STATISTICAL MECHANICS.

[B.C.]

Mechanics, classical

The science dealing with the description of the positions of objects in space under the action of forces as a function of time. Some of the laws of mechanics were recognized at least as early as the time of Archimedes (287?–212 B.C.). In 1638, Galileo stated some of the fundamental concepts of mechanics and, in 1687, Isaac Newton published his *Principia*, which presents the basic laws of motion, the law of gravitation, the theory of tides, and the theory of the solar system. This monumental work and the writings of J. D'Alembert, J. L. Lagrange, P. S. Laplace, and others in the eighteenth century are recognized as classic works in the field of mechanics. Jointly they serve as the base of the broad field of study known as classical mechanics, or Newtonian mechanics. This field does not encompass the more recent developments in mechanics such as statistical, relativistic, or quantum mechanics.

The general principles of classical mechanics are stated in mathematical form. With mathematical logic, one can deduce countless possible motions of bodies and then compare the predictions with experimental observations. Classical mechanics illustrates the essential nature of a physical theory, and it is usually an important ingredient in or a starting point for the various branches of modern physics. Its study offers one the opportunity to become acquainted with mathematical techniques and procedures which are useful in other fields.

In the broad sense, classical mechanics includes the study of motions of gases, liquids, and solids, but more commonly it is taken to refer only to solids. In the restricted reference to solids, classical mechanics is subdivided into statics, kinematics, and dynamics. Statics considers the action of forces that produce equilibrium or rest; kinematics deals with the description of motion without concern for the causes of motion; and dynamics involves the study of the motions of bodies under the actions of forces upon them. An important example of a force whose effect on bodies is often studied is the earth's gravitational force. For some of the more important areas of classical mechanics, *see* BALLISTICS, EXTERIOR; CELESTIAL MECHANICS; COLLISION; DYNAMICS; ENERGY; FORCE; GRAVITATION; KINEMATICS; LAGRANGE'S EQUATIONS; MASS; MOTION; PRECESSION; RIGID-BODY DYNAMICS; STATICS; WORK.

[N.S.G.]

Mechanisms

An assembly of movable parts, having one part fixed with respect to a frame of reference, and designed to produce an effect. Mechanisms are combinations of such moving members as links, gears, cams, belts, chains, and springs held in a rigid frame.

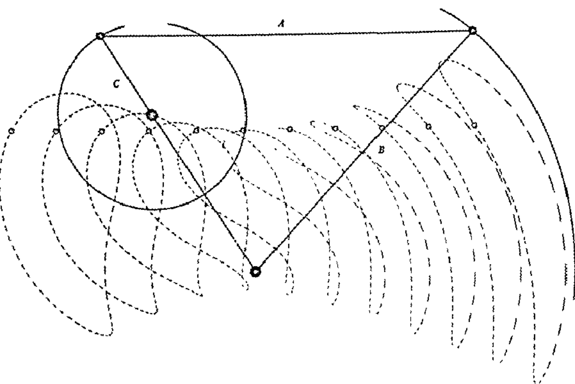
A mechanism may be designed primarily (1) to transmit power greatly in excess of that required to overcome the frictional and dynamic requirements of the mechanism itself, or (2) to produce a desired relative movement of its parts. An example of the first is the slider-crank mechanism in a reciprocating internal combustion engine. An example of the second is the mechanism, consisting of a link, gear sector, and pinion, that converts the movement of a pressure-sensitive tube in a pressure gage to an indication on the dial of the gage.

A machine is an assemblage of one or more mechanisms whose primary purpose is to transmit and control mechanical energy, that is, to do work. Thus, the internal combustion engine is a machine, as are a punch press and a candy wrapper.

An instrument is an assemblage of one or more mechanisms whose primary purpose is to produce controlled relative movements of its parts. The pressure gage, mentioned above, is an instrument. A timepiece and a recording thermometer are instruments.

There are many other devices that are called neither machines nor instruments, but which can be classified in the categories (1) and (2) above. For example, a door latch is a mechanism of the first category; a camera shutter is a mechanism of the second category.

The designer of a machine, instrument, or device contrives first to combine elements to produce desired motions. Next each element is designed to



Paths for crank and rocker mechanisms for A = 4.0, B = 3.5, and C = 2.0. (*From J. A. Hrones and G. L. Nelson, Analysis of Four-Bar Linkage, Wiley, 1951*)

meet requirements of strength, stiffness, and economical manufacture, operation, and maintenance. The two phases of design are inseparable in practice; however, the study of mechanisms is concerned with the first phase, while the design of elements belongs to machine design.

Kinematics of machines. The basis of the modern study of mechanisms, usually called kinematics of machines, is the work of Franz Reuleaux (1829–1905). Reuleaux originated the concept of the kinematic chain, or series of links, in which relative displacements and velocities are independent of which member is fixed. He stated the idea of inversion of a mechanism, whereby, for purposes of analysis, various links of a mechanism are fixed in turn enabling the recognition of significant similarities and differences that are useful in design. Reuleaux was able to show that many inventors of rotary steam engines, in attempting to overcome the disadvantages of reciprocating engines, were in fact merely employing inversions of the slider-crank mechanism. Reuleaux also stated the idea of the instant center.

Analysis. The analysis of relative displacements, velocities, and accelerations occupied the major attention of scholars until recent years (*see* ACCELERATION ANALYSIS; VELOCITY ANALYSIS). Graphical analysis has been favored by most designers because it encourages visualization of a mechanism not yet created, and because it is frequently less cumbersome than a mathematical treatment.

Synthesis. The classical approach to the design of a mechanism relies upon a designer's experience and familiarity with mechanisms of many kinds. To enlarge his repertory of ideas that may be adapted to a specific problem, the designer observes constantly the characteristics of new mechanisms that come within his view. F. D. Jones and H. L. Horton, among others, have catalogued mechanisms for producing various movements. The idea that it may be possible to reverse the process of analysis—that is, to base the kinematic design of a mechanism upon distinguishable principles, rather than to rely solely upon experience and an intuitive sense of fitness—was recognized by Reuleaux, but the problem was not attacked systematically until the 1940s.

Modern synthesis rests upon a study of a family of mechanisms. There are presently two main lines of endeavor: one graphical, one analytical. Graphical kinematic synthesis is particularly useful for mechanisms that perform complex intermittent motions. Mathematical synthesis relies upon equations relating displacement, velocity, acceleration, and time. The increasing availability of large-scale computers has encouraged mathematical studies.

J. A. Hrones and G. L. Nelson have made a systematic graphical approach to synthesis by plotting a large number of paths for crank and rocker mechanisms of various proportions, as illustrated. From such detailed information, a designer can determine the dimensions of a mechanism that w

produce a desired motion. A. S. Hall has prepared a tabular version of similar scope Adjustable-link *point path generators* have also been built to facilitate the search for a suitable linkage.

An analytical approach to synthesis originated about 1940 in Russia and Germany. The work of N. Rosenauer, lately published in English, applies complex variables to the quantitative synthesis of the four-bar linkage. In United States, many papers by F. Freudenstein and others have appeared since 1954, particularly in *ASME Transactions*.

[E.S.F.]

Bibliography: J. S. Beggs, *Mechanism*, 1955; A. H. Church (ed.), *Guillet's Kinematics of Machines*, 5th ed., 1950; J. A. Hrones and G. L. Nelson, *Analysis of the Four-Bar Linkage*, 1951; F. D. Jones and H. L. Horton (eds.), *Ingenious Mechanisms for Designers and Inventors*, 3 vols., 1930–1951; F. Reuleaux, *Kinematics of Machinery*, 1876.

Mechanoreceptors

Receptors provide the organism with information about its environment In higher animals, they are actually the only means by which information of the surroundings is gained and by which reactions to environmental changes are started. Mechanoreceptors provide the organism with information about such mechanical changes in the environment as movement, tension, and pressure. *See* SENSATION.

Receptor-effector mechanisms. In primitive forms of life, such as unicellular organisms and sponges, the receptor and effector functions are built into the same cell (Fig. 1*a* and *b*). The cell is directly excited and reacts to the environmental stimulus. At higher levels of organization, the receptor separates from the effector and appears specialized in function and structure in the receptor cell of the primitive nerve system of actinians (sea anemones). Stimuli of the environment excite the receptor cell, and this excitation is conveyed to another specialized cell, the effector (Fig. 1*c*). In another step of phylogenetic development, a third element appears, the motor nerve cell, which serves as a link between the receptor and effector. This arrangement appears clearly in coelenterates and is carried through, in its essential features, up to the vertebrates (Fig. 1*d* and *e*). The primitive receptor cell may remain at the periphery, as for example in the olfactory receptors of the vertebrates; or it may be buried inside the organism and send protoplasmic branches out to the periphery. A branch of this type, a dendrite or an axon, terminates in its simplest form as a nude ending at the periphery. In more complicated forms, in Pacinian and Meissner corpuscles, the ending is surrounded by accessory tissue of varying structural organization; or the ending appears to terminate on a so-called secondary sense cell, as in certain gustatory, auditory, and visual receptors (Fig. 2). But in all receptors known, the nerve ending is invariably present, however complex the accessory structure around it (Fig. 3). In the two mechanoreceptors whose fine structure has been examined under the electron microscope (the Pacinian corpuscle and the muscle spindle), the nerve ending is nude, lacking not only a sheath of myelin but one of Schwann cells (Fig. 4).

Survival of an organism depends in large part upon its ability to react to changes of the environment. As organized life climbs up the phylogenetic scale, it develops and subsists within increasing limits of environmental change. In general, the greater the organizational level of the animal, the greater are the environmental changes to which it must react and the more varied in form and in degree of functional organization become its receptor organs. Receptors for mechanical changes of the environment (mechanoreceptors) had to develop at an early stage of phylogeny. They were necessary for the avoidance of mechanical objects, for the sensing of water waves transmitted from other animals in the sea, for orientation in space and for maintenance of a position with respect to gravity. Examples are the lateral-line receptors, and the later development of the labyrinth receptors. The development of life on land brought

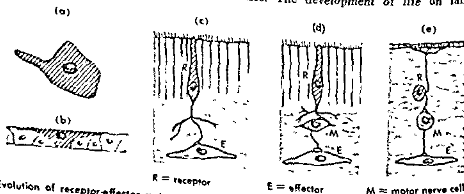


Fig 1 Evolution of receptor-effector systems. (a) Receptor-effector cell of unicellular organism (ameba); (b) receptor-effector cell of multicellular organism (sponge); (c) receptor and effector cells of actinians (tentacle of sea anemone); (d) receptor, motor nerve, and effector cells of coelenterates; (e) receptor, motor nerve, and effector cells in vertebrates. (*b, c, d,* adapted from G. H. Parker, *The Elementary Nervous System*, Lippincott, 1919)

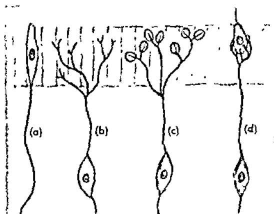


Fig. 2. Differentiation of receptors. (a) Primary receptor cell at the periphery. (b) Nerve cell buried inside the organism sending out branches which terminate as nude receptor endings. (c) Encapsulated receptor endings. (d) Secondary receptor cell

development of mechanoreceptors sensitive to sound vibration of the air. With the growth of internal organs with specialized functions, and particularly, with the development of fast regulatory mechanisms for these functions arose the necessity for receptors (proprioreceptors) sensitive to mechanical stimuli originating inside the organs. Thus, in the vertebrates, mechanoreceptors appear in all organs in which passive or active movements occur, feeding information into the nerve system about movements, tension, or pressure. Mechanoreceptors appear in the digestive tract, in the lungs, in the blood vessels, in the heart, and in the skeletal musculature. The simple mechanoreceptors of the integument of primitive invertebrates are found as bare nerve endings in all animals including the higher vertebrates; they are ubiquitous in the aforementioned organs and in the skin. In addition, around some nerve endings special enclosures of varying complexity are developed in the skin and other structures of the higher vertebrates (Fig. 3). *See* KINESTHETIC SENSATION.

MECHANOELECTRIC CONVERSION

Mechanoreceptors are excited by mechanical disturbances of their surroundings through deformation of their structure, through pressure or tension, or through a combination of these. Whether there is specificity between mechanical stimulus and receptor in excitation is not clear. This question must await further study of the mechanisms of energy conversion at the receptor level; but, in general, the energy requirements for mechanical stimuli to cause a detectable excitation in mechanoreceptors are very low. For example, the receptor membrane of Pacinian corpuscles of the cat's mesentery gives detectable generator potentials with displacement amplitude of its capsule of 10^{-5} cm. *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY.

The transducer sequence. From a physical point of view, mechanoreceptors are energy transducers; they convert mechanical into electrical energy, which in turn triggers the nerve impulse. The mechanisms of nerve-impulse production have been studied most extensively in the muscle spindle of frog by B. Katz; in the stretch receptor of Crustacea by S. W. Kuffler and his coworkers; and in the Pacinian corpuscle by W. R. Loewenstein and his coworkers; and by J. A. B. Gray and M. Sato. In the three receptors studied, deformation

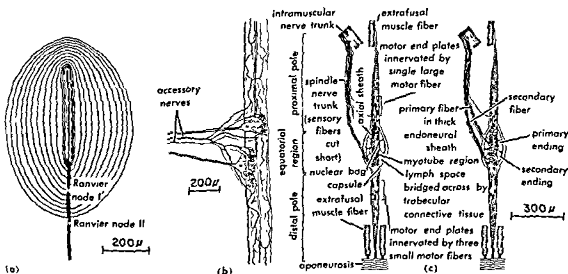


Fig. 3. Schemes of three typical mechanoreceptors. (a) Pacinian corpuscle of the cat's mesentery and skin. (b) Stretch receptor nerve cells of the abdominal musculature of lobster (*Homarus vulgaris*) with dendritic endings embedded in muscle strands (*from J. S. Alexandrowicz, Quart. J. Microscop. Sci., 92:163–199, pt. 2, 1951*). (c) Muscle spindle of the rabbit (*from D. Barker, Quart. J. Microscop. Sci., 89(2):143–186, 1948*).

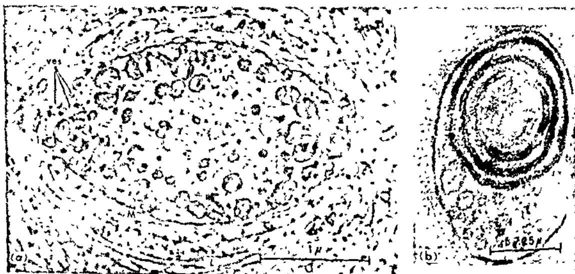
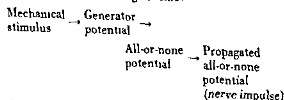


Fig. 4 Fine structure of mechanoreceptors. (a) Electron micrograph of nerve ending of Pacinian corpuscle in the cat, containing vesicular formations (ves), and abundant mitochondria (K) near the ending's membrane (M); L, lamellae of the inner core (from D. C. Pease and T. A. Quilliam, *J. Biophys. and Biochem. Cytol.*, 3(3):331–342, 1957). (b) Electron micrograph of the nerve ending of a muscle spindle (frog) (from J. D. Robertson, *Biochem. Soc. Symposia No. 16*, 3–43, 1959).

leads to a sequence of events which may be summarized by the following scheme:

Mechanical stimulus → Generator potential → All-or-none potential → Propagated all-or-none potential (*nerve impulse*)



The generator potential is the earliest detectable

potential is its graded *nature*; its amplitude increases progressively, without visible steps, if the stimulus strength is progressively increased. When the generator potential reaches a certain critical amplitude, an all-or-nothing potential is discharged in the sense organ which may then propagate as an all-or-nothing nerve impulse along the afferent axon of the receptor (Fig. 5).

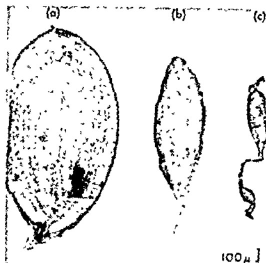
Bioelectric generator and nerve-impulse discharge sites. Generator and all-or-nothing potentials arise at distinctly different membrane sites. The capsular tissue around the nerve ending, constituting *more than 99.9%* of the total mass of a Pacinian corpuscle, does not partake actively in the transducer chain outlined previously (Fig. 6). In Fig. 6, as the capsule of the Pacinian corpuscle was removed (Fig. 6a-c) the preparation was stimulated with two successive mechanical stimuli, the first producing an all-or-none potential, the second (subthreshold) a generator potential. In Fig. 6c mechanosensitivity remains unimpaired in spite of the removal of more than 99.99% *of the* corpuscle. In Fig. 6d and e, although the preparation consisted only of a partly isolated ending, it continued to be as mechanosensitive as before. This may be applicable also to other mechanoreceptors whose nerve ending, like that of Pacinian corpuscles, lies enclosed in a special adventitious tissue. The structure which converts mechanical into electrical energy, that is, the receptor proper, appears to be within the membrane of the nonmyelinated ending. The all-or-nothing nerve impulse, on the other hand, was shown to arise at the first Ranvier node adjacent to the ending (Fig. 7). In Fig. 7c the first Ranvier node is compressed selectively with a fine hook so as to block its activity. Compression of myelinated regions other than the first node does not alter substantially the recorded all-or-nothing impulse.

RECEPTOR MEMBRANE PROPERTIES

Local excitation and the generator elements. The following picture of nerve impulse generation has emerged from work on Pacinian corpuscles by W. R. Loewenstein and collaborators. There is a



Fig. 5. Generation of nerve impulses. Several responses have been superposed on successive oscilloscope sweeps. Note the firing of impulses from a critical level. (a) A crustacean stretch receptor (*from C. Eyzaguirre and S. W. Kuffler, J. Gen. Physiol.*, 39(1):87–119, 1955). (b) In a Pacinian corpuscle (cat) (*from J. A. B. Gray and M. Sato, J. Physiol.*, 122(3):610–636, 1953).



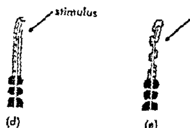




Fig. 6. Localization of receptor membrane (a) Capsule of intact Pacinian corpuscle. (b) Progressive removal of capsule. (c) Only ending and surrounding inner core remain; calibration: 1 msec, 25 μv (d, e) Partial destruction of remaining inner core; calibration: 1 msec. (From W. R. Loewenstein and R. Rathkamp, J Gen. Physiol. 41(6):1245–1265, 1958)

resting potential across the receptor membrane of the nonmyelinated nerve ending. Deformation of the membrane leads to a reduction in membrane potential (depolarization). The depolarization is confined to that region of membrane which has been stimulated mechanically, but does not propagate actively to adjacent regions of the same membrane by local-circuit excitation (Fig. 8, upper diagram of Fig. 9). In Fig. 8 the two microelectrodes are 350 μ apart on the nonmyelinated ending of decapsulated Pacinian corpuscle. The first Ranvier node is grounded and insulated from the ending When the piezoelectric crystal stylus mechanically stimulates the ending near one of the microelectrodes the generator potential recorded from that point has a higher amplitude and faster response than the response from the second microelectrode. The response recorded with the second microelectrode decreases exponentially with the distance from the stimulating stylus, indicating that the active response is restricted to the region of the receptor membrane stimulated mechanically. Current will flow between a depolarized and a normally polarized nonexcited region of the membrane. It is this current which, recorded between an excited and nonexcited region, constitutes the generator current. In opposition to a membrane with all-or-nothing responsiveness, such as the Ranvier node adjacent to the ending or the membrane of a skeletal muscle fiber, excitation in the receptor membrane is not brought about by local flow of current, but only by mechanical excitation. In its general behavior, the receptor membrane of mechanoreceptors resembles the membrane of the motor endplate of skeletal musculature and postsynaptic structures of nerve cells which lack regenerative mechanisms and which seem to have retained within their membrane functional properties of the primitive receptor cell.

For analytical purposes, the receptor membrane may be regarded as composed of functionally independent units or generator elements (Fig. 9). In Fig 9 membrane capacitance and resistance are assumed to be uniformly distributed over the receptor membrane. The transmembrane potential of a generator element is assumed to drop to a fraction of its resting value when the element is acti-

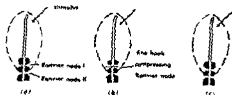


Fig. 7. The site of origin of the nerve impulse. (a, b) Stimulation of decapsulated Pacinian corpuscle by two successive stimuli. Two all-or-none impulses elicited. (c) First Ranvier node blocked, abolishing the all-or-nothing impulse; graded generator potential remains.

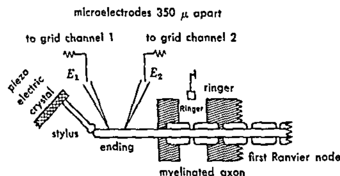


Fig. 8. Local excitation in the receptor membrane. Nonmyelinated ending of Pacinian corpuscle stimulated mechanically by a piezoelectric crystal stylus. Active response is restricted to the region of the receptor membrane stimulated mechanically. Calibration 200 μv; 1 msec. (From W. Loewenstein, *Nature*, 183(4677):1724–1925, 1959)

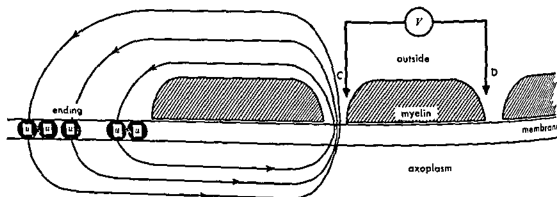


Fig. 9. Electrical model of a receptor. The generator potential (*V*) is measured between the first (C) and second (D) nodes of Ranvier, represented by an equivalent network. (From W. Loewenstein, *Ann. N.Y. Acad Sci.*, *81*:367–387, *pt. 3*, 1959)

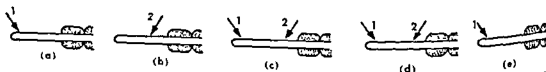


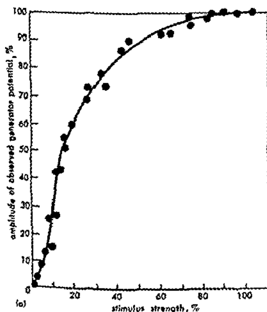
Fig 10. Spatial summation of electric activity in receptor membrane. 1 and 2 are styli of two independent crystal stimulators. Middle and lower beam signal stimuli of styli 1 and 2, respectively. Upper beam for generator potentials, as recorded with a pair of electrodes placed on the myelinated axon. (a, b) Single generator potentials produced when stylus 1 or 2 delivers one single stimulus to its respective site. (c) Generator potential summation produced when styli 1 and 2 both deliver stimulus simultaneously. (d) Summation at a 1-msec delay between stimuli. (e) Refractoriness 1 msec after generator potential production in region below stylus. (From W. Loewenstein, *Nature*, 183(4677):1724–1726, 1959)

vated mechanically. Each generator element may be excited independently of others and may respond to mechanical stimulation with a (active) drop of membrane potential which is confined to that element alone. This feature of local excitation is represented by switching the shunt resistor(r_a) of the active generator into the circuit. Spatial summation of electric activity along the surface of the receptor membrane is an immediate consequence of this mode of excitation (Fig. 10). Summation of excited generator elements may then account for the graded behavior of the generator potential with respect to stimulus strength: as the stimulus strength is progressively increased, a progressively increasing number of generator elements becomes excited. Hence, the generator potential, which is a function of the number of excited generator elements, increases progressively with stimulus strength (Fig. 11). The stepless increase in generator potential would result even if the electrical output of each generator element behaved in an all-or-nothing manner with respect to stimulus strength. *See* CAPACITANCE; RESISTANCE, ELECTRICAL.

The generator potential is built up by a statistically fluctuating number of active generator elements. A mechanical stimulus of a given strength excites a statistical population out of a larger finite number of excitable generator elements. The probability for excitation of generator elements increases as a function of stimulus strength. This mode of excitation appears to have the following consequences: (1) the amplitudes of the generator potential in response to stimuli of standard strength fluctuate at random around a mean value; and (2) when two mechanical stimuli are applied in rapid succession, the number of generator elements excitable by the second stimulus is inversely related to the number of elements excited by the first stimulus.

Refractoriness. The production of a generator potential leaves a refractory condition in the receptor membrane. The refractoriness consists of a reduction in responsiveness of the membrane, which in Pacinian corpuscles lasts 1–3 milliseconds (msec). Thus, a generator response to a standard stimulus which falls on the refractory trail of a preceding conditioning generator potential is smaller than normal, its amplitude being inversely related to the time elapsed after the conditioning generator potential. In addition to time, and quite distinct from a structure with all-or-nothing responsiveness, refractoriness of the receptor membrane is related to the strength of the conditioning stimulus. The greater the conditioning stimulus, the smaller is, within certain limits, the refractory generator potential (Figs. 12 and 13). In Fig. 13 the amplitude of G_2 is measured while G_1 (the total amount of charge transferred in G_1) is varied in *A* and *A'* by varying stimulus strength at constant membrane potential, and in *B* and *B'* by varying the resting potential at constant stimulus strength.

This peculiar behavior of the refractory receptor membrane may be explained as a consequence of the statistical mode of excitation of generator



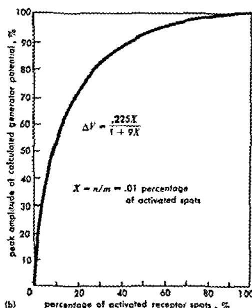


Fig. 11. Input-output relation of receptor membrane. (a) The observed generator potential of the receptor membrane of a typical Pacinian corpuscle as a function of stimulus strength. (b) The calculated generator potential (ΔV) as a function of the number of active generator elements as computed from the receptor analog of Fig. 9. (From W. Loewenstein, Ann. N.Y. Acad. Sci., 81:367–387, pt. 3, 1959)

elements in the receptor membrane. The refractory state is not related to the amount of charge transferred during the conditioning generator potential but to a factor depending upon conditioning stimulus strength.

Receptor inactivation. After repetitive stimulation at high frequency, the responsiveness of the receptor membrane of Pacinian corpuscles becomes markedly reduced (Fig. 14). The effect is fully reversible and has low energy requirements. It reflects a transient inactivation in the receptor membrane which is unrelated to changes in resting membrane potential. The inactivation increases with frequency, (train) duration, and strength of repetitive stimuli. As with the refractory state following a single stimulus, inactivation is not directly related to the amount of charge transferred during repetitive generator potential activity but is related to strength of repetitive stimuli.

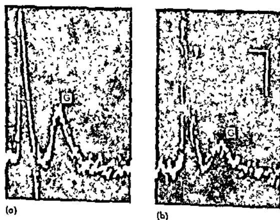


Fig 12 Time factor of refractoriness of receptor membrane. (a) All-or-nothing and generator potential (*G*) caused by two successive mechanical stimuli, 2 *msec apart.* (*b*) *Stimuli 1.5 msec apart, smaller generator potential caused by refractoriness of receptor membrane. (From W. R. Loewenstein and R. Altamirano-Orrego, J Gen. Physiol., 41(4)·805–824, 1958)*

Receptor potentiation. With other, more critical parameters of repetitive stimulation an increase in membrane potential may ensue. Thus, during the period of hyperpolarization (of the order of 10 sec) a generator potential in response to a standard test stimulus is augmented (Fig. 15).

THE PRODUCTION OF NERVE IMPULSES

The triggering mechanism. The nerve impulse of Pacinian corpuscles arises at the first Ranvier node, a structure endowed with typical all-or-nothing responsiveness. When an outward current is passed through it, its resting potential is diminished. As in other excitable tissues with regenerative mechanisms, an all-or-nothing potential is discharged when the resting potential has been lowered to a critical level. Node and receptor membrane of the ending are separated by a cylinder

Fig 13. Size factor of refractoriness Two mechanical stimuli S_1 and S_2 (inset) are successively applied so that the generator potential G_2 in response to the second stimulus falls on the refractory trail of the generator potential G_1 in response to the first. (From W. R. Loewenstein and N. Ishiko, J. Gen. Physiol., vol 42, 1959)

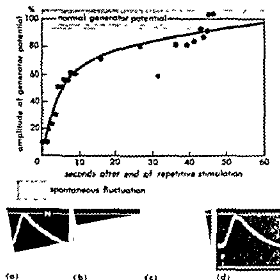


Fig. 14. After-effects of activity in receptor membrane: desensitization. (a) Generator response to a standard test stimulus of the fully rested receptor membrane (Pacinian corpuscle); (b) 0.2 sec, (c) 5 sec, (d) 50 sec after applying a train of repetitive mechanical stimuli at 500/sec for 30 sec to the receptor membrane. Calibration 20 μv; 0.5 msec; (e) recovery of depressed generator potential as a function of time. (*From W. R. Loewenstein and S. Cohen, J. Gen Physiol., 43:347, 1959*)

of myelin of about 250 μ length. There is, however, continuity *for electrical current flow between* both sites through an external path, the interlamellar fluid of the capsule, and through an internal path, the axoplasm. It is thus clear that when the receptor membrane is depolarized by mechanical stimulation, current will flow between active generator elements and the inactive node, outward though the latter (Fig. 9, top). If this current is intense enough to lower the nodal resting potential to the critical *firing level*, an *all-or-nothing potential will be discharged* there which may then propagate as a nerve impulse along the axon towards the nerve centers. This is the way nerve impulses are produced at receptors and how information of the outside world is conveyed to the nerve centers.

Code of sensory messages. Because the conductive elements of the nerve systems of higher organisms, *the axons, along which the nerve impulse must propagate*, behave in an all-or-none manner, there is no possibility for gradation in amplitude of nerve impulses. Nature has endowed organisms with a rather stereotyped information system in which messages are sent in a code of equal all-or-nothing pulses. Thus, the only way of conveying quantitative information in nerve messages in a given information channel is by changing the sequence of its staccato content.

It has been established that the generator current which flows between the active receptor membrane and the first node increases progressively with *strength of stimuli*; and that the rate of rise of a generator potential increases with amplitude of the generator potential, or more directly, with the difference in potential between node and receptor membrane. Thus, with suprathreshold stimuli of different strength, the resulting generator currents will cause the nodal membrane potential to drop to the critical level at a rate directly related to stimulus strength. Hence the frequency of discharged impulses at the node will be a function of the rate of rise of the generator potential and indirectly of the stimulus strength (Fig. 16). This is then one way to feed messages of different quantitative content into the nerve system. The greater the stimulus intensity, the higher the impulse frequency of the message. The uncovering of this code is due chiefly to Lord Adrian.

In animals of higher organization, receptors occur in groups. A pull on a muscle excites not one, but many spindle receptors. A weight applied to the skin excites a group of Pacinian corpuscles. If one single receptor out of a group were to be killed, it would make little or no difference to the total message reaching the central nerve cells. This leads to another manner by which the content of a message may be graded. *The greater the deformation of an organ, the greater is the probability for excitation of different homonymous receptors.* Thus, as a weight applied to the skin is increased, a *larger skin surface is deformed, a larger number of* receptors are excited, and hence a larger number of parallel axons become involved in impulse con-

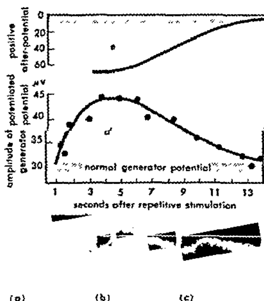
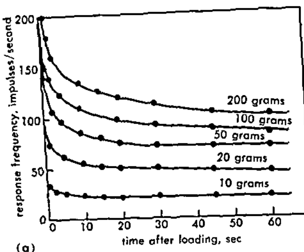


Fig. 15. After-effects of activity in receptor membrane: potentiation. Responsiveness of receptor (Pacinian corpuscle) is tested (a) with standard stimulus under *fully rested conditions; (b) 5 sec; (c) 18 sec* after a train of repetitive stimuli at 500/sec for 5 sec. (d) Time course of resulting potentiation; (e) hyperpolarization. (*From W. Loewenstein and S. Cohen, J. Gen. Physiol., 43:347, 1959*)



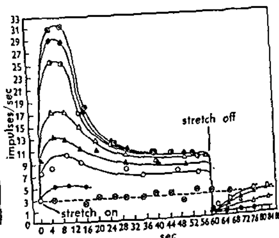


Fig. 16. Impulse frequency as a function of stimulus strength. (a) Frog muscle spindle at various tensions (*from B. H. C. Matthews, J. Physiol., 78(1):1–53, 1933*).

(b) Pressure receptor of frog skin at various degrees of stretch (*from W. R. Loewenstein, J. Physiol., 133 (3).588–602, 1956*).

duction. Besides, since generally several receptor endings are twigs of one axon, there will be considerable mixing of impulses into the common axon. Consequently, because the number of receptors activated increases with strength, the frequency of total impulses in the common axon may also increase by this mechanism. Further mixing may still occur because of convergence of several axons onto nerve cells along the way to the centers.

There are, thus, essentially two ways by which the central nervous system may analyze the content of a receptor message: by counting the number of parallel information channels engaged in impulse traffic, and by gaging the frequency and sequence pattern of impulses transmitted in each information channel.

RECEPTOR ADAPTATION

Adaptation in impulse discharge. One of the striking properties of mechanoreceptors is their ability to adapt to stimuli. Adaptation is best shown by applying a single mechanical stimulus of a given suprathreshold strength to a receptor organ and maintaining it so indefinitely without change. The receptor may respond with an initial burst of impulses at a given frequency; but then, in spite of the fact that the stimulus is continuously sustained, the frequency of impulses diminishes and may in some receptors fall to nought. The receptor is said to have adapted to the stimulus. The rate of adaptation varies greatly for different mechanoreceptors. Thus, a Pacinian corpuscle adapts to zero impulses within a few milliseconds; other skin receptors take seconds; and stretch receptors of the skin and of the muscle spindle maintain stationary discharges even after minutes of steady stimulation. It is in the latter type that the stimulus strength–impulse frequency relationship is most readily seen (Fig. 16).

Adaptation in generator potential. The mechanism of adaptation is as yet poorly understood. However, adaptation depends, at least in part, upon the ability of the receptor organ to sustain a constant generator potential. In fast-adapting receptors, such as the Pacinian corpuscle, the generator potential decays spontaneously to zero within 4–7 msec in the face of a continuously sustained suprathreshold stimulus. In slowly adapting or nonadapting receptors, such as the muscle spindle and the slow-stretch receptor of Crustacea, the generator potential is well sustained for some time above the firing threshold of the node (Fig 17). In the former case impulses can obviously be produced only during the 1–2 msec in which the generator potential is above firing threshold; whereas in the latter, a continuously sustained barrage of impulses may result.

The recognition of the fact that adaptation depends upon the receptor organ's ability to sustain a generator potential moves the question of adaptation one stage ahead of the node, to the receptor membrane. But a formulation of the question in terms of the receptor membrane's capacity to sustain a generator potential should not be made. Too little is known about mechanical properties of receptors to be certain that when a stimulus is

(a) (b)

Fig. 17. Adaption of generator potential. Generator potential in response to a mechanical stimulus is applied and sustained continuously for about 4 sec. (a) In a fast-adapting Pacinian corpuscle; calibration, 1 msec between dots (*from W. R. Loewenstein, J. Gen. Physiol., 41(4).825–845, 1958*). (b) In a slowly adapting crustacean stretch receptor; calibration 1 sec (*from C. Eyzaguirre and S. W. Kuffler, J. Gen. Physiol., 39(1).87–119, 1955*).

applied continuously to a whole receptor organ the stimulus which reaches the receptor membrane proper is also maintained. On the contrary, the evidence available indicates that in the fast-adapting Pacinian corpuscle, in certain cutaneous touch receptors, and in the fast-stretch receptor cell of Crustacea, only the initial phasic component of deformation of the receptor organ also deforms the receptor membrane proper; whereas the protracted static component of deformation appears to be entirely absorbed by the accessory tissue around the receptor membrane. An hypothesis which considers a mechanical component in receptor adaptation has been formulated.

EFFERENT CONTROL OF RECEPTORS

Modulation at the receptor level. Receptor inhibition and facilitation are discussed in the following sections.

Receptor inhibition. The impulse discharge of certain mechanoreceptors can be modulated by centrifugal impulses. An example of this is the stretch receptor cell of Crustacea. This cell receives accessory nerve fibers which end on the receptor dendrites. A. W. Kuffler and collaborators have shown that the effect of impulses conveyed by nerve fibers to the receptor cell is to clamp its membrane potential near resting value by increasing apparently specifically potassium ion (K^+) conductance. In this manner, mechanically elicited generator potentials *are prevented from reaching* the critical level for impulse firing, and sensory discharges in response to stretch are promptly suppressed at the receptor level (Fig. 18a). In Fig. 18a the inhibitory axon to the receptor is stimulated (between arrows), preventing the generator potential from reaching firing level.

Receptor facilitation. Certain tactile receptors of the frog's skin can be modulated by centrifugal impulses from the sympathetic system. Loewenstein has shown that arrival of sympathetic impulses or application of sympathomimetic substance increases the sensitivity of touch receptors to mechanical stimuli (Fig. 18b). In Fig. 18b there is an excitation of new receptor units, indicating an increase in excitability; also, there are repetitive discharges of some units during sympathetic stimulation, indicating lowering of the adaptation rate. *The sympathetic effect is mediated by an adrenalin-like substance* which appears to diffuse from the sympathetic terminal to the touch-receptor ending. The anatomical relations between the afferent and efferent systems here involved are not known.

Modulation at the accessory structure. A muscle spindle in the rabbit consists schematically of a bundle of muscle fibers (intrafusal muscle fibers) whose central portion (nuclear bag) beds the nude receptor ending (the receptor proper) of the thick myelinated afferent axon (Fig. 3). The intrafusal muscle receives, in addition, a set of motor fibers whose impulses cause it to contract. The motor control of the spindle has been studied extensively by R. Granit and collaborators and by Hunt and Kuffler. Stretching out the central nuclear bag excites the receptor ending, causing impulses to travel centripetally along the thick axon. The in-

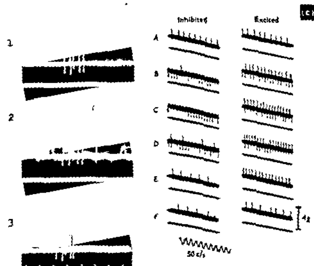


Fig. 18. Efferent control of receptors. (a) Intracellular recording from a crustacean stretch receptor submitted to constant stretch (from C. Eyzaguirre and S. W. Kuffler, J. Gen. Physiol., 39(1):155–184, 1955). (b) Sensitization of touch receptors (frog) by centrifugal sympathetic impulses. Impulse discharge of a group of touch receptors in response to a standard mechanical stimulus: 1, before, 2, during, 3, after arrival of sympathetic impulses to the skin (from W. R. Loewenstein, J. Physiol., 132(1):40–60, 1956). (c) Supraspinal control of muscle spindle (cat soleus). Impulse discharge at a standard stretch. A, E, F, before and B, C, D, after stimulation of inhibitory loci (left) and excitatory loci (right) of the contralateral inferior colliculus (from E. Eldred, R. Granit, and P. A. Merton, J. Physiol., 122(3):498–523, 1953).

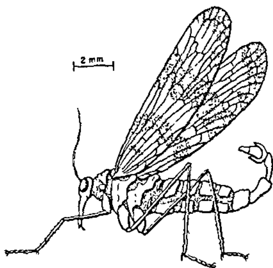
trafusal muscle fibers lie in parallel with the large muscle fiber of the skeletal muscle in which they are cribbed. Thus, shortening of the skeletal muscle during contraction will cause the intrafusal fiber to slacken and thereby the afferent impulse discharge to cease. If, under these conditions, motor impulses are conveyed to the intrafusal fibers, the latter appear to contract at a relatively fixed length (the ends of the intrafusal fiber are inserted in the endomysium or perimysium of the skeletal muscle). Thus the slack is taken up, the receptor ending is stretched, and impulses are fired again to the centers The motor innervation of the intrafusal bundle appears thus to control the responsiveness of the sense organ by setting the tension of its accessory structure. *See* NERVOUS SYSTEM; NERVOUS SYSTEM (INVERTEBRATE).

[W.R.LE.]

Bibliography: E. D. Adrian, *The Physical Background of Perception*, 1947; H. Davis, Biophysics and physiology of the inner ear, *Physiol. Revs.*, 37(1):1–49, 1957; R. Granit, *Receptors and Sensory Perception*, 1955; J. A B. Gray and M. Sato, Properties of the receptor potential in Pacinian corpuscles, *J. Physiol.*, 122(3):610–636, 1953; H K. Hartline, H. G Wagner, and E. F. MacNichol, Jr., The peripheral origin of nervous activity in the visual system, *Cold Spring Harbor Symposia Quant. Biol.*, 17:125–141, 1952; B. Katz, Depolarization of sensory terminals and the initiation of impulses in the muscle spindle, *J Physiol.*, 3:261–282, 1950; S. W. Kuffler, Synaptic inhibitory mechanisms, properties of dendrites and problems of excitation in isolated sensory nerve cells, *Exptl. Cell Research*, 5(suppl.):493–519, 1958; W. R. Loewenstein, Excitation and changes in adaptation by stretch of mechanoreceptors, *J. Physiol.*, 133(3): 588–602, 1956; W. R. Loewenstein, The generation of electric activity in a nerve ending, *Ann. N.Y. Acad. Sci.*, 81:367–387, pt. 3, 1959; W. R Loewenstein and S. Cohen, Post-tetanic potentiation and depression of generator potentials in a single nonmyelinated nerve ending, *J. Gen. Physiol.*, 43:347–376, 1959, W. R. Loewenstein and N. Ishiko, Properties of a receptor membrane, *Nature*, 183(4677): 1724–1726, 1959; W. R. Loewenstein and R. Rathkamp. The sites for mechano-electric conversion in a Pacinian corpuscle, *J. Gen. Physiol.*, 41(6): 1245–1265, 1958; W. R. Loewenstein, S. J. Socolar, and S. Cohen, on the kinetics of inactivation in a receptor membrane, *J. Gen. Physiol.*, vol. 43, 1959; G. Svaetichin, G. and E. F. MacNichol, Jr., Retinal mechanism for chromatic and achromatic vision, *Ann. N.Y. Acad. Sci.*, 74:385, 1958.

Mecoptera

A small order of insects called the scorpionflies. Characteristic of the adult insect is the peculiar prolongation of the head into a beak, which bears chewing mouthparts. They are small to medium in size. The insects either have two pairs of large, net-veined wings of equal size, often with dark



Male scorpionfly, order Mecoptera, genus *Panorpa*.

areas, or have short and aborted wings. The legs are long and slender. In some species, the male abdomen has a terminal enlargement which is held [illegible] resembles a scor- [illegible] re eruci- [illegible] upae ex- [illegible] arate; that is, the appendages are free and used in locomotion. The Mecoptera are found in moist habitats within densely wooded areas. The adults are omnivorous but feed chiefly on small insects.

[B.E.R.]

See INSECTA.

Medicine

The term medicine has two general meanings. The first indicates any material, notably in the drug category, which is given to prevent, alleviate, beneficially alter, or stop a disease process. In a broader sense, medicine denotes the field of science devoted to healing. Many subdivisions exist and more ramifications appear almost daily. Included in the area of medicine are the clinical specialties of surgery, pediatrics, psychiatry, obstetrics, and others. Internal medicine is the specialization which deals with internal diseases of a nonsurgical nature.

Related to the clinical specialties, particularly in regard to medical education and research, are the basic medical sciences. These include, among others, anatomy, physiology, pharmacology, biochemistry, and microbiology. Midway between the basic and the clinical sciences lies pathology, the study of the structural and functional alterations caused by diseases or abnormal states. *See* ANATOMY, REGIONAL; BIOCHEMISTRY; MICROBIOLOGY; PATHOLOGY.

An important area in all specialties is preventive [illegible] nedicine [illegible] nmunity. [illegible] atters of [illegible] prevention, mass treatment, and statistical appraisals of health matters. It is also concerned with

socioeconomic factors related to physical and mental well-being. *See* PUBLIC HEALTH.

The mental health of individuals and nations has received much attention since 1930. Prior to that time mental illness was largely a matter of disgrace, secrecy, and custodial care in primitive, prisonlike hospitals. With the recognition of the importance of the dynamics of psychiatry, the advent of better therapeutic measures, particularly specific drugs which promote mental well-being, and the realization that preventive measures are of vital concern to all, the field of mental health today represents one of the largest areas in medicine. *See* ABNORMAL BEHAVIOR.

Much attention has been given to aviation and space medicine, both of which consider the problems peculiar to their fields in the light of clinical and experimental information, as well as that supplied by other sciences and disciplines. *See* AVIATION MEDICINE; SPACE BIOLOGY.

Socialized medicine is that form which exists under the direct control and financing of the state. The National Health Service of Great Britain is the best-known example at the present time, although other systems exist.

Other subdivisions of medicine whose names are largely self-explanatory include veterinary, legal, tropical, and military medicine.

Probably the most familiar aspects of medicine to the layman are the two clinical aspects, diagnosis and therapeutics. Diagnosis includes taking a patient's history, a physical examination, and possibly laboratory and special examinations. Therapeutics is the administration of some form of preventive or remedial treatment to the patient.

Although medicine is based primarily upon scientific information and method, an important feature is the relationship between the physician

Mediterranean Sea

The Mediterranean Sea lies between Europe, Asia Minor, and Africa. It is completely land-locked except for the Strait of Gibraltar, the Bosporus, and the Suez Canal. The Strait of Gibraltar is 13 miles wide at the boundary between the Atlantic and the Mediterranean (Great Europa Point, Gibraltar, to Punta Almina, Ceuta), and the Bosporus is as narrow as ½ mile, widening to 2 miles at the Black Sea boundary (Rumeli Burnu, European Turkey, to Anadolu Burnu, Asiatic Turkey). The Suez Canal, a man-made channel connecting the Mediterranean with the Red Sea, is 87½ miles long, about 300 ft wide, and 37 ft deep. In latitude, the Mediterranean extends from 30°15′N in the Gulf of Sidra to 45°47′N in the Gulf of Trieste; in longitude, from 5°21′W at Great Europa Point to 36°12′E in the Gulf of Alexandretta.

The Mediterranean is conveniently divided into an eastern basin and a western basin, which are joined by the Strait of Sicily and the Strait of Messina. The Strait of Sicily, between Cap Bon, Tunisia, and Sicily, is 77 miles wide, but the Strait of Messina, between Punta Sottile, Sicily, and the Italian mainland, is only 1.7 miles across.

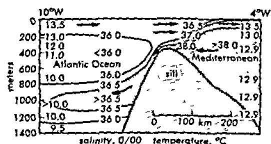
The major subdivisions of the western basin are the Alboran Sea, which is the portion west of Cabo de Gata, Spain, and is named for the Isla de Alboran; the Balearic or Iberian Sea, between Spain and the Islas Baleares; the Ligurian Sea, the gulf north of Corsica; and the Tyrrhenian Sea, lying between Corsica, Sardinia, Sicily, and the Italian mainland. The eastern basin includes the Adriatic Sea, between Italy down to the heel of the boot (Capo Santa Maria de Leuca) and the Balkan Peninsula north of Corfu; the Ionian Sea, which lies south of the Adriatic as far as the southern tips of Sicily and the Greek mainland; the Aegean Sea, the waters north of Kithera, Andikithera, Crete, Karpathos, and Rhodes, as far as the Dardanelles; and the Sea of Marmara, which extends from the Dardanelles to the Bosporus.

The Mediterranean Sea has numerous islands, particularly the Ionian Sea (which has the alternate designation of The Archipelago) and the eastern Tyrrhenian and Adriatic. The largest islands are Sicily, 25,710 km²; Sardinia, 24,040; Cyprus, 9250; Corsica, 8720; and Crete, 8380.

The total water area of the Mediterranean is 2,516,000 km² and its average depth is 1494 m.

The greatest depth in the western basin is 3625 m, in the Tyrrhenian Sea. The eastern basin is deeper, with a greatest depth of 5530 m in the Ionian Sea about 35 miles off the Greek mainland. The Adriatic mainly overlies the continental shelf, but there is a basin in the southern part with a greatest depth of 1315 m. The sill depth in both the Strait of Gibraltar and the Strait of Sicily is about 300 m.

The Atlantic tide disappears in the Strait of Gibraltar. The tides of the Mediterranean are predominantly semidiurnal. The eastern and western basins each has a standing wave system, with



Vertical distribution of salinity and temperature in west-east section through Strait of Gibraltar along 36°N, according to G. Schott, 1944. Arrows show principal spreading of the waters. (From G. Dietrich and K. Kalle, Allgemeine Meereskunde, eine Einführung in die Ozeanographie, Gebrüder Borntraeger, Berlin, 1957)

nodal lines extending roughly from Barcelona to Bougie and from Kerme Korfezi to Kithera to Tobruch. The tide range elsewhere is about 1 ft. At the ends of the Straits of Sicily and Messina, high water occurs in exactly opposite phase, giving rise to strong hydraulic currents, particularly in the narrow Strait of Messina, the legendary seat of Scylla and Charybdis. The Adriatic has a progressive tide, with range as great as 3 ft in the northern end, radiating about an amphidromic point near 43°N, 15°E.

Since evaporation over the whole Mediterranean greatly exceeds the supply of water from precipitation and river runoff, the surface salinity is higher than that of the Atlantic, increasing from a value of less than 36.50 ‰ in the Strait of Gibraltar to over 39.00 ‰ in the extreme east. Surface temperatures reach 27 or 28°C in the southern and eastern gulfs in the summer, elsewhere ranging from 21 to 25°C. In winter, the northern Adriatic may be as cold as 8°C, but off Egypt and Syria the temperature may not drop below 16°C.

Its subsurface waters are all formed within the Mediterranean, owing to the shallow sill separating it from the Atlantic. To a depth of about 600 m, these waters are characterized by a salinity maximum between 38.60 and 39.90 ‰ and temperature of 14°C or greater; they are formed at the surface, where evaporation is most significant. The deeper waters are formed in winter along the northern shores of the sea, where winter cooling has its greatest effect; they are about 1C° colder than the intermediate waters and 0.20 ‰ or 0.30 ‰ lower in salinity. In turn, the deeper water of the eastern basin has a temperature about 0.5C° higher and salinity 0.25 ‰ higher than that of the western basin.

The excess of evaporation within the Mediterranean requires that a surface current of Atlantic water enter the Straits of Gibraltar. The layer of intermediate water, since it has a greater density than Atlantic water at the same depth, gives rise to an outgoing countercurrent which pours down over the continental slope and whose waters can be detected for thousands of miles in the Atlantic. The volumes of water in the currents are some 15 times as great as the actual evaporation; thus, some 1.75×10^6 m³/sec enter the Strait of Gibraltar, or enough to renew the entire 3.759×10^6 km³ in about 70 years. *See* Atlantic Ocean. [J.LY.]

Megaphone

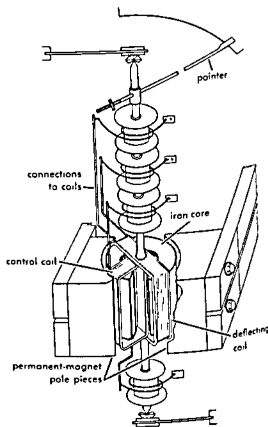
A device for increasing the efficiency of sound radiation and confining the sound of a speaker along a preferred direction. In the configuration most often used, the megaphone consists of a conical horn. The user speaks into the smaller end of the horn, which is of approximately the same cross-sectional area as a man's mouth.

The megaphone performs its functions in several ways. The geometric constriction by the horn increases the effective sound power radiated along the megaphone axis by at least the ratio of 2π to the solid angle subtended by the horn. The small throat area of the horn more effectively matches the acoustic impedance of the speaker's mouth. The large area of the horn mouth has a more nearly resistive acoustic impedance than the smaller area of the speaker's mouth; therefore, the effective power radiated by the horn is greater by roughly the ratio of the two areas.

Functionally, the megaphone acts as an acoustical transformer. The factors which influence its design are mouth diameter, throat diameter, length, and rate of flare. A wave acoustics analysis permits the optimum choice of these parameters in any practical case. *See* Impedance, acoustic; Loudspeaker; Sound. [W.J.C.]

Megger

A high-range ohmmeter commonly used as a portable instrument for measuring the high resistance of electrical insulating materials of the order of 20,000 megohms at 1,000 volts. The Megger is a crossed-coil ohmmeter, which is a permanent magnet, moving-coil, direct-reading instrument whose accuracy is independent of the exact voltage applied to the specimen. The source of voltage, usually contained in the same case as the ohmmeter itself, may be a hand-cranked or motor-driven generator, a plug-in power supply, or a battery.



Megger coils and magnets. (*J. G. Biddle Co.*)

Equipment is also made for testing up to 10,000 volts.

The moving-coil system and the magnets of the Megger are shown in the illustration. The torque developed by the deflecting coil is proportional to the current in the test specimen. The torque of the control coil is proportional to the voltage applied to the specimen. Coils A and B are connected to develop opposing torques, and the pointer therefore takes a position where the two torques are balanced, which is the indicated resistance of the test specimen. See INSULATION RESISTANCE TESTING.

[C.E.A.]

Meiosis

A special type of cell division which occurs in diploid (or polyploid) tissues and brings about a reduction in chromosome number. It is, therefore, the antithesis of syngamy. See CELL DIVISION; SYNGAMY.

The essential factor in both plant and animal fertilization is fusion of the nuclei of paternal and maternal gametes; also the nuclei of a particular species must possess a definite and constant number of chromosomes. From this it follows that there must also be a mechanism of compensation which would provide for a reduction of chromosome number at some period during the life cycle of a sexually reproducing organism to offset the doubling brought about by syngamy. The egg and sperm contribute an equal number of like chromosomes in forming the zygote. In man, for example, each gamete contributes 23 chromosomes to make 46 in the zygote, or 23 pairs. Since the members of each pair are similar in genetic content to each other they are referred to as homologous chromosomes, or homologs. Each parent, therefore, contributes a single, or haploid, set of chromosomes through its gamete, to give a double, or diploid, set in the zygote and all cells derived from the zygote by mitosis.

Meiosis, in its basic features, consists of two nuclear divisions in rapid succession, but with only one longitudinal division of the chromosomes. The first division involves the separation of homologs that have paired with each other, and two nuclei or cells are formed which divide again to produce four haploid cells. Accompanied by appropriate divisions of the cytoplasm, sexual gametes in animals, or asexual spores in plants are formed. The term maturation division is synonymous with meiosis.

Animals. In animals the meiotic divisions producing sperm in the testis, or male reproductive organ, occur during spermatogenesis. The meiotic divisions producing eggs in the ovary, or female reproductive organ, occur during oogenesis.

Spermatogenesis. Each testis is divided into lobes, in each of which are groups of cells called spermatogonia. These enlarge to become the primary spermatocytes. The first meiotic division divides each of these into two secondary spermatocytes, while the second division divides the latter into spermatids. Thus four spermatids are formed from each primary spermatocyte. A process known as spermiogenesis transforms the spermatids into motile spermatozoa or sperm. See SPERMATOGENESIS.

Oogenesis. The ovaries of animals vary widely in form, but certain of the cells become oogonia, which enlarge to become primary oocytes, usually containing large amounts of reserve food. The first division separates the homologs as in spermatogenesis, but an unequal division of the cytoplasm buds off a small nucleated cell, the polar body, leaving the secondary oocyte virtually unchanged in size. The latter divides again and a second polar body is extruded. During this time the first polar body may have divided, and these two plus the second polar body result in three small cells clustered on the surface of the now haploid egg, where they eventually disintegrate. See OOGENESIS.

Plants. The process by which spores are formed in plants is known as sporogenesis. The time relationships between meiosis and syngamy vary in plants, and the meiotic products are haploid asexual spores rather than gametes. In some algae and fungi, meiosis follows rather than precedes syngamy, and the resultant spores germinate and grow to produce a haploid thallus. This structure, the gametophyte, produces gametes by mitosis, and the zygote resulting from their union undergoes meiosis without further growth. There is, consequently, no diploid body.

Higher plants have a similar life cycle except that the diploid zygote develops into a diploid sporophyte. This structure produces a sporogenous tissue, the cells of which undergo meiosis and form megaspores (female) and microspores (male). The two processes are called megasporogenesis and microsporogenesis, respectively. The megaspore develops into a megagametophyte which forms eggs by mitotic divisions; the microspore forms a microgametophyte which eventually produces sperm. See FLOWER (BOTANY).

Stages of meiosis. A number of different stages in meiosis are recognized.

Leptonema. This is the earliest recognizable prophase stage. It does not differ appreciably from a similar stage in mitosis except that the cells and nuclei are generally quite large. The chromosomes are long and thin and are differentiated along their length into beadlike structures called chromomeres. These are constant in size, number, and position, a fact which enables the cytologist to make use of them as landmarks for the recognition of particular chromosomes. In the lily, for example, it has been estimated that the diploid set of chromosomes may contain as many as 2000 chromomeres.

Zygonema. Zygonema is the stage when the homologs begin to unite actively in pairs. Pairing, or synapsis, may begin at any place along the length of the chromosome, but so exact is the union that homologous chromomere pairs with homologous

chromomere. When there is a physical interruption in chromomere sequence, there is an interruption in pairing. Eventually a zipperlike action brings the homologous chromosomes into close alignment along their entire length.

Pachynema. The term refers to the more obvious thickness of the chromosomes, brought about by the fact that the chromosomes exist in pairs and that they have contracted somewhat. If zygonema is the active stage of pairing, pachynema may be considered the stable stage; it may also be of some duration. The chromosomes appear to be present in the haploid state, but this is merely because the two homologs of each pair are closely appressed.

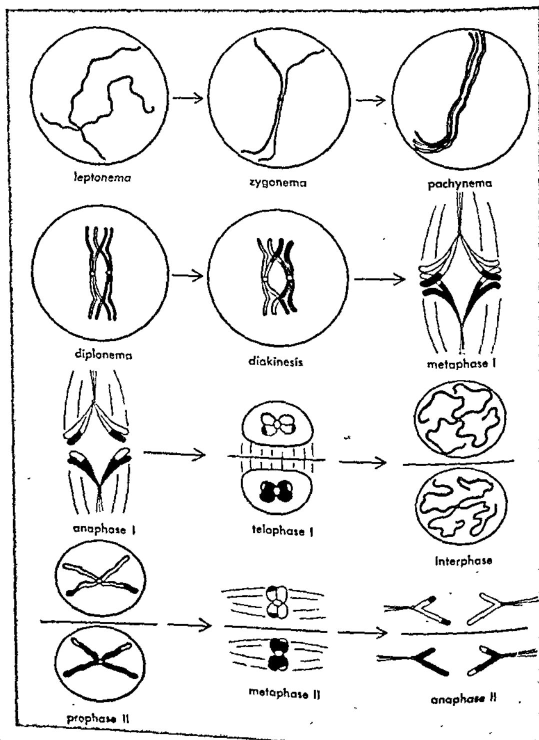


Diagram of the principal stages of meiosis. (*From M. M. Rhoades, Meiosis in maize, J. Heredity, 41:59–67, 1950*)

The nucleolus, formed by one pair, is very evident at this stage. Each pair is called a bivalent.

Diplonema. This stage is initiated by a longitudinal separation of the paired chromosomes. The lapse of the pairing force reveals that each chromosome is longitudinally double; each bivalent therefore consists of four chromatids. The term tetrad has been used to describe this structure. As the pairing force wanes, the homologs separate from each other. The separation, however, is generally incomplete, and one or more points of contact are maintained. If only one contact is present, the bivalent resembles a cross or a rod; if two, a closed loop; if three or more, a series of connected loops. These points of contact are called chiasmata, and in clear preparations it can be seen that two of the chromatids, one from each homolog, have crossed over. Since the two sister chromatids of each chromosome do not separate laterally from each other, and since the crossover is between nonsister chromatids, the chiasma is the only point of exchange that keeps the pair of chromosomes together.

Diakinesis. This stage is characterized by a pronounced contraction of the chromosomes, the disappearance or detachment of the nucleolus from its associated bivalent, the disappearance of the nuclear membrane, and the even distribution of the bivalents throughout the cell.

Metaphase I. The spindle has formed by this time, and the bivalents are oriented with their centromeres directed towards the poles. The chromosomes are even more contracted, and in clear microscopic preparations it can be seen that contraction has been due to the development of a coil within the chromosome; each chromosome, in fact, closely resembles a wire door spring. A difference between mitotic and meiotic chromosomes should be noted. In mitosis, each chromosome lies with its divided centromere on the metaphase plate. In meiosis, the bivalent has two functionally undivided centromeres which are oriented in the long axis of the spindle, with the distance between centromeres regulated by the position of the proximal chiasma. There appears to be an active repulsion between homologous centromeres which keeps them apart and oriented.

Anaphase I. The movement of the chromosomes from the metaphase plate to the poles constitutes anaphase. The movement involves a shortening of the spindle fibers connecting the chromosomes to the poles and an elongation of the spindle elements between the two moving groups of chromosomes. Since the separation is one of homologs, each anaphase group is made up of a haploid set of chromosomes instead of a diploid set of chromatids, as in mitosis.

Telophase I and interphase I. These stages, which may be lacking in some organisms, involve the regrouping of the chromosomes into an organized nucleus, with a relaxation of the coiled structure.

Second meiotic division. The second meiotic division differs from a normal mitotic division only in that the chromosomes are longer and thinner, and they are present in a haploid instead of a diploid number. Genetically, however, the separating chromatids need not be similar since chiasma formation may have altered the genic combinations. If the first meiotic division is looked upon as a reduction division, the second is an equational one, but those terms are applicable only to chromosomes and not necessarily to the genes contained therein. The end result of the second meiotic division is four haploid nuclei, and the reduction in chromosome number has been accomplished.

Significance of meiosis. In relation to sexual reproduction, meiosis is the mechanism for the production of gametes and for the reduction in chromosome number; it is coexistent with syngamy in reproduction of all sexually breeding animals. In plants, however, meiosis separates the alternation between sporophyte and gametophyte generations; only indirectly is it involved in sexuality.

Significance to genetics. Meiosis provides the physical mechanism for the segregation and independent assortment of genes, and through chiasma formation it leads to the breaking up of linked allelic combinations. Chiasma formation is, therefore, the cytological equivalent of genetical crossing over. See CHROMOSOME; GENE. [C.P.SW.]

Meissner effect

A magnetic property of superconductors such that any magnetic field is excluded from their interiors. The effect was discovered by W. Meissner in 1933. No magnetic flux can penetrate a superconductor when it is placed in a magnetic field. Moreover, if a superconducting material is cooled below its transition temperature while in a magnetic field, that portion of the field within the material will be expelled. For a discussion of this and other properties of superconductors *see* SUPERCONDUCTIVITY.

A superconducting material is perfectly diamagnetic. This magnetic property is fundamental to superconductors and not a consequence of their perfect conductivity, which would assure only that any interior magnetic field would not vary with time. *See* DIAMAGNETISM.

The complete Meissner effect is observed only in superconductors of high chemical and physical purity. The effect is incomplete in most superconducting alloys and metals containing internal strains or dissolved impurities. Such materials may have no electrical resistance and yet expel only a fraction of any magnetic field contained within. This incomplete Meissner effect is attributed to small internal regions remaining normally conducting during the superconducting transition while the surrounding material becomes superconducting. Magnetic flux is then threaded through and "trapped" in these regions, preventing their ever becoming superconducting. Although the superconducting regions may be linked so as to give rise to a perfectly conducting path, the impure material as a whole may appear to be only partially diamagnetic. [M.D.F.]

Melanconiales

An order of fungi of the class Fungi Imperfecti. A great number of species, for example *Gloeosporium* and *Colletotrichum*, are plant pathogens commonly causing anthracnose (dark or black, limited lesions). There are about 90 genera and 1000 species recognized Conidiophores, or hyphae bearing conidia, are densely aggregated on a cushion-shaped or disk-shaped mass of hyphae, an acervulus. Acervuli, at first covered by the epiderm or the bark, form a pustule and later burst through the substratum, that is, are erumpent. The conidia are always slimy spores. This order contains only one family, Melanconiaceae. The distinction between Sphaeropsidales and Melanconiales is often difficult.

The genera of the Fungi Imperfecti are usually arranged into spore groups depending on characteristics of the spore, such as the number of cells in the spore, shape of the spore, and whether the spore is bright or dark. [illegible] cussed below.

1. *Gloeosporium*. Members of the 150 known species are parasites on leaves, herbaceous stems, or fruits. Many species are conidial stages of *Glomerella* or *Gnomonia* (genera of the Ascomycetes). The acervulus, a fruit body containing conidia, is disk-shaped, pale pink or fuscous in hue; it is originally subepidermal or subcuticular but later bursts through the subsurface. The conidia are oblong or ovate. *G. concentricum* causes leaf spot of cabbage. *G. fructigenum*, causing bitter rot of apples and pears, is a conidial stage of *Glomerella cingulata*. *G. nervisequum*, the plane tree scorch, is a conidial stage of *Gnomonia veneta*.

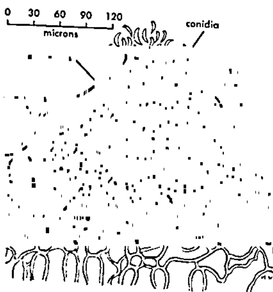


Fig 1. Acervulus and conidia of *Gloeosporidiella ribis*. (After H. Klebahn, 1906)

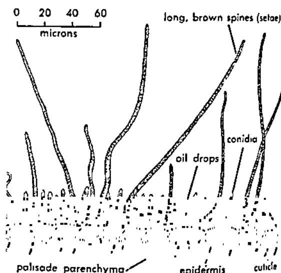


Fig. 2. Acervulus of *Colletotrichum primulae* with conidia and setae. (After G. Moesz, 1942)

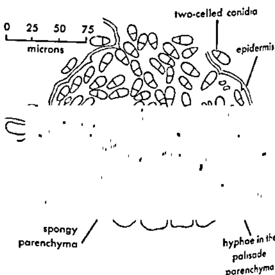


Fig. 3. Acervulus and conidia of *Marssonia populi* (After H. Klebahn, 1918)

2. *Gloeosporidiella* is similar to *Gloeosporium* but the conidia are strongly curved. *G. ribis* (Fig. 1), the leaf spot of currant and gooseberry, is the conidial stage of *Drepanopeziza* (*Pseudopeziza*) *ribis*.

3. *Colleotothrichum* is similar to *Gloeosporium* but its conidiophores are interspersed with long, black, irregularly occurring spines (Fig. 2). The 70 species are parasites and imperfect stages of *Glomerella*. *C. lindemuthianum*, an anthracnose of bean, may belong to *Glomerella lindemuthianum*. *C. linicola* causes the seedling blight of flax. *C. gloeosporioides*, producing wither-tip of citrus plants, is a conidial stage of *Glomerella cingulata*.

4. *Myxosporium* is also similar to *Gloeosporium*, but the acervuli are formed in the bark of trees.

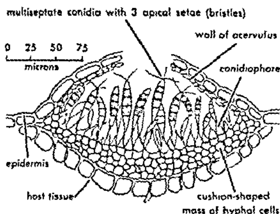


Fig. 4. Acervulus and conidia of *Pestalotia theae*. (After E. J. Butler, 1918)

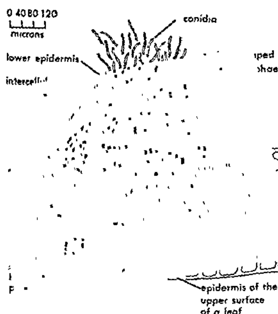


Fig. 5. Acervulus and conidia of *Cylindrosporium padi*. (After V. B. Stewart, 1914)

The 60 species are mostly saprophytes or weak parasites. *M. lanceola* is found on oak twigs.

5. *Cryptosporiopsis* differs from *Myxosporium* only in the oblong or cylindrical shape of the conidia. Species are conidial stages of *Pezicula*. *C. (Myxomysporium) corticola*, the apple surface canker, is the imperfect stage of *Pezicula corticola* (ascomycete).

6. *Sphaceloma* is like *Gloeosporium*, but the conidiophores arise from a pseudoparenchymatous stroma, sometimes appearing as a sporodochium. The 50 species are imperfect stages of *Elsinoë; S. rosarum* is the cause of leaf spot of roses, and *S. fawcettii*, of citrus scab.

The important genus of the group Phaeosporae which has one-celled, dark spores, is *Melanconium*, mostly conidial stages of *Melanconis*. *M. fuligineum* causes bitter rot of grape.

A genus of the Hyalodidymae group which has two-celled hyaline spores is *Marssonia*. The conidia are often unequally two-celled (Fig. 3). The 70 species are parasitic, usually causing leaf spots, and are often the imperfect stage of Discomycetes. *M. rosae*, the black spot of roses, is a stage of *Diplocarpon rosae*.

A genus of the group Phaeophragmiae which has dark spores containing two or more cells with cross septae, is *Pestalotia (Pestalozzia)*. The 50 species are mainly parasitic. The conidia are usually three-celled, the central cells brown, and the end cells hyaline. The apical cell has two or more setae or spines. *P. theae* causes gray blight of tea. *P. palmarum* causes spotting of coconut leaves (Fig. 4). *P. guepini* is found on leaves of *Camellia*.

A genus of the Hyaloscolecosporae, a group which has bright threadlike spores, is *Cylindrosporium*. The 50 species are typically parasitic on leaves. Several species are conidial stages of *Coccomyces (Higginsia)*. *C. padi* (Fig. 5), which causes "shot-hole" disease of cherry leaves, is a stage of *Coccomyces hiemalis*.

A genus of the group Phaeostaurosporae which has dark, one-celled spores, starlike in form, is *Asterosporium*, whose conidia are starlike, four-armed, and with each arm septate. *A. hoffmanni* is a saprophyte on twigs of beech. *See* FUNGI; PLANT DISEASE. [N.F.B.]

Melanterite

A mineral having composition $FeSO_4 \cdot 7H_2O$. Melanterite occurs mainly in green, fibrous or concretionary masses or in short, monoclinic, prismatic crystals. Fracture is conchoidal and luster is vitreous. Hardness is 2 on Mohs scale and specific gravity is 1.90. The mineral is similar to chalcanthite, has an astringent taste, and is readily soluble in water.

Melanterite is a common secondary mineral derived from oxidation and hydration of iron sulfide minerals such as pyrite and marcasite. Its occurrence is widespread. It is not an ore mineral.

Melanterite can be crystallized from aqueous solution at temperatures up to 56°C. The mineral is unstable in dry air at room temperature, however. It readily dehydrates into a yellowish powder. [E.C.T.C.]

Melilite

A mineral sorosilicate with complex composition crystallizing in the tetragonal system. Melilite forms short prismatic crystals with a poor basal cleavage. The hardness is 5 on Mohs scale, and the specific gravity varies from 2.95 to 3.04, increasing with increasing amounts of aluminum. The luster is vitreous to resinous and the color is white, yellow, greenish, reddish, or brown. Melilite is the name given to the complete solid-solution series extending from akermanite, $Ca_2MgSi_2O_7$, to

gehlenite, $Ca_2Al_2SiO_7$. Aluminum substitutes in the structure for magnesium and also for silicon to maintain electrical balance. The gehlenite end of the series is more common; pure akermanite is found only in slags. Melilite is a rock-forming mineral found in recent volcanic rocks which have formed from magmas low in silica, as in central Italy, particularly at Vesuvius. Gehlenite is found in limestones at the contacts of intrusive rocks. *See* SILICATE MINERALS. [C.S.HU.]

Melioidosis

A disease of rodents which, rarely, is transmitted to man. It is caused by a bacterium, *Pseudomonas pseudomallei* (*Malleomyces pseudomallei*). Melioidosis is a grave malady, existing mostly in southeastern Asia, which is seldom diagnosed before death. Two cases have been reported in the Western Hemisphere.



Ultraphotomicrograph of *Pseudomonas pseudomallei.*

Pseudomonas pseudomallei is a short, motile, aerobic, gram-negative, granular rod which is similar to *Pseudomonas mallei*. Infection is chronic in wild rats and the organisms are discharged in urine and feces. Rat fleas and *Aedes aegypti* mosquitoes can harbor the organism.

Melioidosis in man occurs in two forms, a pulmonary septicemic form, which ends fatally in 3 days to 3 weeks, and a rare chronic form lasting months or years. Cultural methods afford the best means of diagnosis. Chloromycetin is the most effective drug. For taxonomy *see* PSEUDOMONADACEAE; *see also* CHLORAMPHENICOL. [W.R.MI.]

Bibliography: G. S. Wilson and A. A. Miles, *Topley and Wilson's Principles of Bacteriology and Immunity*, Vol. 2, 4th ed., 1955.

Melon growing

The production of edible fruits of the watermelon (*Citrullus vulgaris*) and the muskmelon (*Cucumis melo*), including cantaloupes, honeydew melons, and Persian melons. Botanically, the melon fruit is a pepo (kind of berry); the matured ovary and its seeds are completely surrounded by an extensive development of flower tissues which form the rind. The seeds are attached to the ovary wall. In most forms there is no obvious demarcation between the outermost rind (exocarp) of the ovary and the inner layers of edible tissues.

Characteristics. Both watermelons and musk melons are frost-tender, trailing annuals of the family Cucurbitaceae, order Campanulales. Male and female flowers are borne on the same plant. The males begin to appear before the females; the latter are easily distinguished by the tiny undeveloped melon subtending the corolla. The pollen is sticky and heavy and is carried to the female flowers by insects.

Varieties within a species intercross readily, but muskmelons and watermelons will not cross with each other or with the cucumber, squash, or pumpkin. When grown for seed for planting, varieties within a species must be separated by ¼ mile or more to minimize intervarietal crossing. Although intervarietal crossing does not affect the properties of the fruit in the generation in which the cross occurs, the seeds from such crosses will produce mixed plants and fruits, maybe of inferior quality. *See* CAMPANULALES; MUSKMELON; WATERMELON. [V.R.B]

Diseases of melons. Various diseases of muskmelons cause an annual loss of about $10,000,000 on the 325,000 acres planted in the western, southwestern, midwestern, and mid-Atlantic regions of the United States. Diseases also cause a $3,000,000 loss on watermelons produced on about 500,000 acres in the southwestern, southern, and southeastern states.

Muskmelon diseases. Losses from seed decay and seedling blight resulting from infection by the soil fungi *Pythium irregulare, P. ultimum,* and *Rhizoctonia solani* can be reduced by treating 100 lb of seed with 5 oz of thiram together with shallow sowing in warm, moist soil.

Muskmelon vines wilt, leaves yellow and die, and fruits wither and sunburn when affected by root rot and crown blight, incited by such soil fungi as *Fusarium solani* and *Pythium* spp. These diseases can be partially controlled by crop rotation, treat

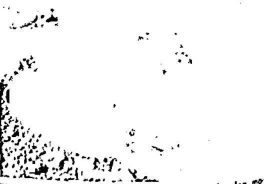


Fig. 1. Cantaloupe leaf affected by muskmelon mosaic; vein banding on first leaf results from seed transmission of the virus.

ing seed with mercuric chloride, maintaining good nitrogen fertility, and avoiding soil moisture tensions of more than 60 centibars. Powdery mildew causes the greatest damage in the warm, dry, rainfree growing season of the Southwest. The fungus *Erysiphe cichoracearum* first appears as white spots on the underside of the older leaves. The spots later coalesce, and affected leaves lose their normal dark-green color, become pale, then brown and then shrivel. Fruits on infected vines ripen prematurely and lack the desired texture, flavor, and sugar content. This disease is controlled on susceptible varieties by dusting them with 1% karathane. Powdery-mildew-resistant cantaloupes No. 5 or 45, and Georgia 47 should be grown where conditions permit.

The leaves of young plants infected with the muskmelon mosaic virus show conspicuous dark-green bands along the larger veins, (Fig. 1); later leaves usually do not show vein banding, but display a severe to mild yellow-and-green mottling and may become distorted and curled. The symptoms of crown blight are accentuated when plants are also virus-infected. The mild mosaic patterns that sometimes occur on immature fruits disappear with normal development, but the soluble solids are low and flavor is poor in infected fruits. The muskmelon mosaic virus is commonly seed-transmitted, ranging from more than 95% in freshly extracted seed to less than 5% in 3-year-old seed. It is readily transmitted by aphids. Use of virus-free seed, weed-free culture, and insect control delays the occurrence of the virus and minimizes crop losses.

Watermelon diseases. Watermelons are seriously affected by a vascular wilt incited by the fungus *Fusarium oxysporum* f. *niveum* which is able to live in the soil for many years, even in the absence of its host. If very young plants are attacked, the seedlings may rot before or after emergence from the soil, or the plants may become stunted. When mature plants are attacked, the tips of the runners wilt, and gradually the entire plant shrivels and dies. The woody part of the stem is brown and the



Fig. 2. (a) Watermelon anthracnose showing sunken decay spots on fruit. (b) Enlarged view showing detail of anthracnose lesions on watermelon fruit and fruiting bodies of *Colletotrichum lagenarium*, the causal fungus. (Photos courtesy D. E. Ellis)

roots exhibit cankers. Infected plants produce few fruits and these are of poor quality. The disease is controlled by planting resistant varieties, such as Blacklee, Charleston Gray, Dixie Queen, Hawkesbury, Klondike R-7, Leesburg, and Peacock. Anthracnose, incited by the seed-borne fungus *Colletotrichum lagenarium*, causes black, angular spots on leaves and sunken, circular lesions on fruits (Fig. 2). Infected melons are unmarketable and typically decay in transit and storage. Keeping land free from cucurbits for 3 or more years, use of disease-free western-produced seed, or treatment of seed for 5 min with 1-1000 mercuric chloride, and regular applications of sprays or dusts containing copper or zineb and ziram throughout the growing season will reduce the incidence of the disease. *See* PLANT DISEASE; PLANT VIRUS.

[J.T.M.]

Bibliography: A. Stefferud (ed.), *Plant Diseases*, USDA Yearbook Agr., 1953.

Melting point

The temperature at which a solid changes to a liquid. For pure substances, the melting or fusion process occurs at a single temperature, the temperature rise with addition of heat being arrested until melting is complete. The direct transition from solid phase to gas phase is not properly called melting, but preferably sublimation.

Melting points reported in the literature, unless specifically stated otherwise, have usually been measured under an applied pressure of 1 atm, usually 1 atm of air. (The solubility of air in the liquid is a complicating factor in precision measurements.) The pressure dependence of the absolute melting temperature T_m is given by the Clapeyron equation

$$\frac{dT_m}{dP} = \frac{T_m \, \Delta V_f}{\Delta H_f}$$

where ΔV_f is the change in volume, ΔH_f the heat absorbed during the fusion process, and P is the applied pressure. Upon melting, all substances absorb heat, and most substances expand; consequently an increase of pressure will normally raise the melting point. A few substances, of which water is the most notable example, contract upon melting; thus, the application of pressure to ice at 0°C will cause it to melt. Large changes in pressure are required to produce significant shifts in the melting point; thus, for example, a pressure of 10 atm lowers the melting point of ice by only 0.075°C.

A sufficient decrease in temperature at ordinary pressures will cause all pure substances except helium to freeze to solids; the lowest normal melting point is that of hydrogen at 14°K, and one of the highest is that of rhenium at 3700°K. Liquid helium can be transformed into a solid only by applying a pressure in excess of 25 atm.

For solutions of two or more components, the melting process normally occurs over a range of

temperatures, and a distinction is made between the melting point, the temperature at which the first trace of liquid appears, and the freezing point, the higher temperature at which the last trace of solid disappears, or equivalently, if one is cooling rather than heating, the temperature at which the first trace of solid appears (*see* EQUILIBRIUM, PHASE). Measurement of the freezing point of a solution and the difference between it and the freezing point of the pure solvent provides a convenient method of determining the molecular weight of a dissolved solute, because the freezing point of a solution is lower than that of a pure solvent *See* SOLUTION; SUBLIMATION; TRIPLE POINT. [R.L S.]

Memory

Memory is evident in behavior when, after a passage of time without practice, past events are recalled or previously learned skills are performed. It is generally measured in terms of the amount of the material which is remembered or the amount which is forgotten, these measures being the converse of one another. Students of memory have been concerned with the effect on the amount remembered of such variables as the kind of material to be recalled, the methods of presenting the material, the delay between presentation and recall, the type of events occurring during this delay period, and the methods used to elicit recall. Another area of investigation has been concerned with the question of what goes on in the brain during the process of remembering and forgetting, where in the brain these processes occur, and the effect on memory of such factors as brain injury, drugs, oxygen deprivation, and radiation.

Methods of studying memory. In 1855 Hermann Ebbinghaus, a German psychologist, published the results of several years of research on memory. These studies were a significant break-through for they demonstrated that the higher mental processes, like the psychophysical processes such as the judgment of just noticeable differences in sensation, were amenable to measurement. This enabled research on memory to depart from the subjective and philosophical methods which had dominated discussions of the topic. In addition, it represented the invention of a method which was to dominate research on memory for many years. This method involved two significant innovations. First, it provided a rigidly defined memory task, the nonsense syllable list consisting of items which had little if any meaning for the subject and which were unrelated to one another. Second, it provided a method of quantification. The number of syllables in the list defined the relative difficulty of the task, and the savings in the number of trials required to relearn a list after a delay defined the amount remembered. With this method and with himself as subject, Ebbinghaus derived the classic curve of forgetting illustrated in Fig. 1. Most studies of remembering and forgetting have been concerned with the effect of a variety of variables on this function.

Materials used in memory experiments. Ebbinghaus invented the nonsense syllable to provide a large series of items of uniform difficulty and lacking in association value. By taking two consonants and a vowel at random and putting them together he produced 2300 syllables such as xef, leb, wuc which he could arrange in lists of different lengths. Workers since Ebbinghaus have revised the lists so as to decrease even further the association between items and to provide lists of more or less association value. In addition, his successors improved on the method of presenting the lists with the use of the memory drum. This instrument (Fig. 2) has appeared in a variety of models, but basically it enables the experimenter to present stimuli one at a time with precise control over stimulus parameters. It consists of a drum on which are affixed the

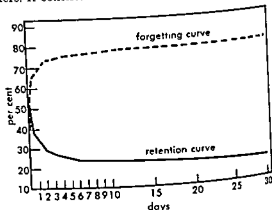


Fig. 1. Curves showing the percentage forgotten or retained at various intervals after learning as determined by comparing the number of trials to relearn a list of nonsense syllables with the number of trials to original learning. (*From data of H. Ebbinghaus, Über das Gedachtnis, Duneker and Humblot, 1885*)

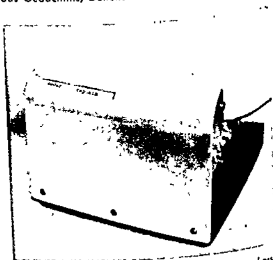


Fig. 2. A modern memory drum showing paired-associate phrases in the aperture. (*Lafayette Instrument Company, Lafayette, Indiana*)

lists to be memorized. As the drum revolves, individual items of a list appear in the aperture. More modern and elaborate devices of this sort utilize motion picture film and projectors to achieve the same purpose, but have the advantage of greater variety and control of experimental procedures.

Next in popularity to the nonsense syllable list has been the paired-associate list. An example of this is the foreign language vocabulary lesson in which the English word and its foreign equivalent are given. Those for experimental purposes are constructed of pairs of words having high or low association value depending on the particular purpose of the study. Lists made up of such pairs as "exact-precise" have high similarity value, while those made up of such pairs as "lawless-distinct" have low similarity value. Although the lists are usually of words, they may be of pictures or other material. As with the nonsense syllables, paired associates are usually presented on a memory drum.

The nonsense-syllable list and the paired-associate list owe some of their popularity to the fact that many similar lists of equal difficulty may be constructed. Such lists thus provide standard materials so that the results of several experiments may be compared. In addition, graded scales are easily constructed simply by increasing the length of the list. But it is the fact that the lists are composed of discrete items making rigor in quantification possible that has made them the material of choice in learning and memory studies. *See* VERBAL LEARNING.

The lists have been criticized, however, for representing too restricted a sample of the substance of memory. Accordingly, other more complex materials have frequently been used. Ebbinghaus had used poetry in several of his studies, using as his measure the number of whole repetitions required for mastery. Many investigators since then have also used poetry, often dividing the poem into many separate units or phrases for purposes of quantification. However, poems often have different meanings and associations for different subjects, making standardization of the material impossible.

Many other materials, both spoken and written, have been used; among these are passages of narrative, number series, photographs, simple geometric and common object pictures, and so forth. However, with this material it is difficult to achieve as precise control over the elements of association value, similarity of items, and uniformity of difficulty as it is with the nonsense or paired-associate material. Consequently, most of the work in human remembering and forgetting has been done with the latter more rigorous material.

Generally, the investigations using the materials discussed thus far have been concerned with long-term memory based on many repetitions of the material to the level of complete mastery. Another area of investigation has been concerned with immediate or short-term memory. Many of the materials listed previously may be used in testing immediate memory; they are then presented only once, and the number of items recalled is used as the measure of immediate memory. More often, however, the digit span or some similar test is used. Here, several numbers are read off to the subject and he is asked to repeat them. The longest list that he can repeat is called his memory span. Since this method can give a quick estimate of immediate memory it is commonly used in the clinic.

Another class of materials which has been used in the study of memory includes tests of motor skill. Most skills are relatively persistent in memory, probably because of the fact that they are originally learned well beyond the level of simple mastery. The skills which have been used most often in this respect include a variety of performance tests such as sending and receiving code messages, guiding a stylus through a maze, following targets, card sorting, and the like.

Although most of the studies in memory have used human subjects, many have used other animals, particularly when brain damage, drugs, or other agents have been involved. The maze for animals is somewhat similar to the studies involving rote memory of a serial list with humans. Here, each turn in the maze is, for a rat, an item, which like the nonsense syllable for human beings, is part of a list which is learned by frequent repetitions. In the typical experiment the animal learns the problem after many repetitions, then, after a delay or experimental treatment such as the infliction of brain damage, it is tested for retention or savings in number of trials to relearn the skill. Highly reliable results can be obtained with this type of test.

Another device is the puzzle box, in which the animal is required to remember the sequence of acts it had previously learned to open a latch. This method is more difficult to score, having fewer discrete items, and is therefore less reliable than the maze. The jumping stand and its modifications has been used most often for testing memory for visual patterns. This is a device on which the animal, usually a rat, stands facing two designs several inches in front of and just beyond a space over which he is encouraged to jump. If, upon jumping, he strikes the design which is correct, it falls over and he lands safely on a platform and is fed. If he strikes the incorrect one, he falls a few inches into a safety net and is not fed. Essentially the animal is forced to choose between two alternatives, one of which he has learned conceals a reward.

The method most often used for testing memory in monkeys is the Wisconsin general test apparatus. A modification of this apparatus appears in Fig. 3. It is essentially a method for presenting the animal with a choice between two or more items. A typical memory study is carried out by showing the animal which of two objects conceals a peanut, lowering the screen for a delay, raising it, and allowing the monkey to choose. This procedure would test short-term memory. Long-term memory

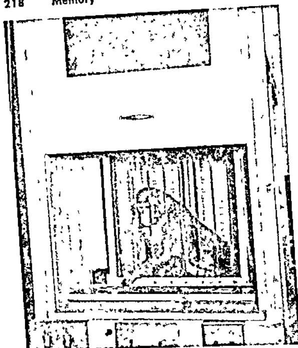


Fig. 3. A modification of the Wisconsin general test apparatus showing a monkey responding to a pattern which he has learned conceals a peanut

would be tested by teaching the animal over a series of trials always to choose the same one of two alternatives to obtain a peanut reward, and after an interval, testing to see if he remembers which of the two is correct. If he fails to recall correctly on the first trial, he would be retrained to see if he could relearn in fewer trials than it took originally. This apparatus has been used to test many aspects of memory in monkeys and other animals. These aspects include the improvement in memory in phylogenetic development, the effects of brain damage, and the effects on memory of variations in stimulus, delay, and response.

Measures of memory. Memory is manifest in four somewhat different ways which, though having many features in common, may be discussed separately. An entire scene or complex of events may be recollected. Such a redintegrated memory, as this type of memory is called, may be aroused by the presence of one or a few of the original cues, as for example when the sound of a voice recalls not only the image of a person but also the circumstances of the dinner at which he was introduced. Although these memories can be very vivid and rich in detail, they are not very satisfactory research material. There is rarely any reliable way of comparing the memory with the original event with respect to accuracy of content or number of details. Thus quantification is not usually possible. Where checking has been possible, it has been found that the memories are contaminated and distorted with the intrusion of more recent events, or with confabulated details.

Much of the research with this type of memory has been concerned with recovering such memories by means of such techniques as electrically stimulating the brain, administering such drugs as sodium amytal, or by having the subject free-associate to various stimuli. A research device frequently used with this type of memory is the testimony experiment. Typically, a scene such as that of a crime is enacted before the subject, and he is later asked to describe the details of the scene. Both the accuracy of his description and the number of items recalled are used as scores.

Another measure of memory is the recognition score. Recognition of an item occurs in everyday experience when an event such as having seen a person, heard a tune, or smelled burning toast is recognized as part of past experience. The characteristic feature of this type of memory is the familiarity surrounding the object, the awareness that the event has occurred before. The typical experiment involves presenting a sample and later asking the subject to identify the sample among a variety of other materials. With monkeys this is usually done by teaching him first that by lifting a certain object, say a toy car, he will obtain food. On the recognition test, the car appears as one of the objects among several others. The monkey is said to have recognized the object if he goes to it first. This test may be given in the apparatus illustrated in Fig. 3. In humans, the experiment is similar. Typically, the subject is shown an item, say a photograph, then later required to select the one he had seen before. The score is the number of items correctly identified. This method has not been popular as a research tool primarily because of the difficulty in presenting the material at the time of response and because of the difficulty in presenting the cue in a context of standard association value.

Recall or retention of some performance learned in the past is evident in everyday experience when a skill such as swimming is performed with little difficulty on the first swim of each summer or when a French word elicits its English equivalent. In laboratory experiments the subject is ordinarily taught some task and later asked to recall it. A monkey, for example, is taught that a peanut is always concealed under the green object of a pair. Recall after a delay or after brain damage or drug administration is demonstrated if, on the first test trial, he goes to the green object. With human subjects, a wider variety of tasks may be employed. The most common method is to teach subjects a list of paired associates using the memory drum. Some time later a test is given in which the first member of the pair appears in the aperture of the memory drum and the subject is asked to say or write the associate. The percentage correct is the retention or recall score. Both because of the ease in presenting materials and availability of many standard lists, this is a very popular research tool. Another type of recall test very popular in the clinic because of its ease of administration is the memory span test described earlier. It is generally used only for immediate recall.

A fourth method of measuring memory is based on the premise that if any residual memory of a learned habit remains, relearning of that habit should take fewer repetitions to achieve the same level of perfection as had previously been attained. The measure is the savings or relearning score; that is, the difference between the number of trials to learn and to relearn. This is one of the most popular methods of measuring memory primarily because very rigorous methods may be applied both to the administration of the test and the treatment of the data.

The memory of many species of animals has been tested on a wide variety of tasks with this method. Rats have been taught mazes of many kinds and the number of trials to learn has been compared with number of trials to relearn after such treatments as simple passage of time, brain damage, electroconvulsive shocks, and stress. In fact much of the knowledge of memory in relation to the brain is based on the rat's memory of maze performance.

An increasing amount of information on memory, perhaps more relevant to human beings, is becom- [illegible] learned through any sense modality can be tested. Thus it has been possible in the subhuman primate to compare memory for visual, auditory, tactual, and olfactory tasks; to localize memory for these tasks in specific structures in the brain; and to study the effect of a variety of factors on memory.

As would be expected, the three measures of memory (recognition, recall, and relearning) vary in sensitivity, yielding scores which differ considerably in the estimate of the amount remembered. Figure 4 shows a comparison of the percentage of nonsense syllables retained over 2 days as determined by each of the methods. It is apparent that any estimate of the amount retained at any period during the interval would differ considerably depending on the method used to measure retention. Nevertheless, the course of forgetting over time is roughly the same regardless of the method of measurement. There appears always to be a rapid initial drop so that within a short time, as in the recall measure, as much as 75% of the material is forgotten. Following this initial effect further loss occurs slowly and is scarcely perceptible in the retention score. These curves, however, are based on serial lists either of nonsense syllables or paired associates. Other conditions of testing yield curves having a very different shape, indicating that under some circumstances the course of forgetting may be quite different from that illustrated in Fig. 4.

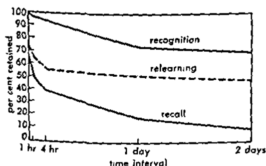


Fig 4. Curves showing the percentage of nonsense syllables retained as measured by three different methods. (After data of C. W. Luh, *The Conditions of Retention*, Psychol. Rev. Co., 1922)

Factors which affect memory. In the preceding sections the principal methods used in studies of memory have been described. The results of these studies indicate that memory is affected by a variety of factors some of which are most easily described in terms of the memory task itself, others in terms of the brain.

Nature of the material. The meaningfulness of material has considerable influence on the degree to which it will be remembered. In fact, Ebbinghaus invented the nonsense syllable in order to remove meaningfulness of material as a variable in his studies. Words or events which have special meaning for the subject, or which may be associated with situations which do, are more easily remembered than those which are comparatively neutral. Similarly, almost any feature which makes an item more vivid or more accentuated, thus commanding the attention of the subject, will make it more easily remembered. This is shown in the type of experiment in which a syllable is very easily remembered because it is inserted in a list of numbers. The position of a word in a list is also a determinant of the ease with which the word can be recalled, items at the beginning and end being better remembered than those in the middle.

Recall also depends on the length of the list to be recalled. This is apparent in immediate recall, where it is evident that as each digit is added to the problem, recall becomes increasingly difficult and is, on the average, impossible for a series of more than seven items. If the list is longer than this, it has to be grouped into seven or so units if it is to be remembered after one presentation. Material which cannot be so broken up requires more than one presentation if it is to be recalled accurately. Longer lists require more presentations.

Many experiments have shown that the affective quality of the material influences the likelihood of its being recalled. The majority of these studies indicate that unpleasant, traumatic, or anxiety-producing materials are poorly retained. It is not clear, however, that this effect is due so much to poor retention of these materials as it is to their being poorly perceived in the first place. Material which is learned just to the level of mastery is not retained as well as material that is overlearned. It is for this reason that repetitive drills are necessary if material is to be retained for a long time. A subject who is set to remember materials and has some expectation of the way in which he is going to be asked to recall the material is better able to do so than one who is not so prepared. These facts indi-

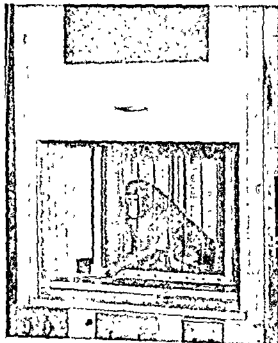


Fig. 3. A modification of the Wisconsin *general test apparatus* showing a monkey responding to a *pattern* which he has learned conceals a peanut.

would be tested by teaching the animal over a series of trials always to choose the same one of two alternatives to obtain a peanut reward, and after an interval, testing to see if he remembers which of the two is correct. If he fails to recall correctly on the first trial, he would be retrained to see if he could relearn in fewer trials than it took originally. This apparatus has been used to test many aspects of memory in monkeys and other animals. These aspects include the improvement in memory in phylogenetic development, the effects of brain damage, and the effects on memory of variations in stimulus, delay, and response.

Measures of memory. Memory is manifest in four somewhat different ways which, though having many features in common, may be discussed separately. An entire scene or complex of events may be recollected. Such a *redintegrated* memory, as this type of memory is called, may be aroused by the presence of one or a few of the original cues, as for example when the sound of a voice recalls not only the image of a person but also the circumstances of the dinner at which he was introduced. Although these memories can be very vivid and rich in detail, they are not very satisfactory research material. There is rarely any reliable way of comparing the memory with the original event with respect to accuracy of content or number of details. Thus quantification is not usually possible. Where checking has been possible, it has been found that the memories are contaminated and distorted with the intrusion of more recent events, or with confabulated details.

Much of the research with this type of memory has been concerned with recovering such memories by means of such techniques as electrically stimulating the brain, administering such drugs as sodium amytal, or by having the subject free-associate to various stimuli. A research device frequently used with this type of memory is the testimony experiment. Typically, a scene such as that of a crime is enacted before the subject, and he is later asked to describe the details of the scene. Both the accuracy of his description and the number of items recalled are used as scores.

Another measure of memory is the *recognition* score. Recognition of an item occurs in everyday experience when an event such as having seen a person, heard a tune, or smelled burning toast is recognized as part of past experience. The characteristic feature of this type of memory is the familiarity surrounding the object, the awareness that the event has occurred before. The typical experiment involves presenting a sample and later asking the subject to identify the sample among a variety of other materials. With monkeys this is usually done by teaching him first that by lifting a certain object, say a toy car, he will obtain food. On the recognition test, the car appears as one of the objects among several others. The monkey is said to have recognized the object if he goes to it first. This test may be given in the apparatus illustrated in Fig. 3. In humans, the experiment is similar. Typically, the subject is shown an item, say a photograph, then later required to select the one he had seen before. The score is the number of items correctly identified. This method has not been popular as a research tool primarily because of the difficulty in presenting the material at the time of response and because of the difficulty in presenting the cue in a context of standard association value.

Recall or retention of some performance learned in the past is evident in everyday experience when a skill such as swimming is performed with little difficulty on the first swim of each summer or when a French word elicits its English equivalent. In laboratory experiments the subject is ordinarily taught some task and later asked to recall it. A monkey, for example, is taught that a peanut is always concealed under the green object of a pair. Recall after a delay or after brain damage or drug [illegible] the first test [illegible] sub- [illegible] ayed. The most common method is to teach subjects a list of paired associates using the memory drum. Some time later a test is given in which the first member of the pair appears in the aperture of the memory drum and the subject is asked to say or write the associate. The percentage correct is the retention or recall score. Both because of the ease in presenting materials and availability of many standard lists, this is a very popular research tool. Another type of recall test very popular in the clinic because of its ease of administration is the memory span test described earlier. It is generally used only for immediate recall.

applied later. The animals shocked early suffered memory loss whereas those shocked later did not, suggesting that in the process of laying down the memory trace in brain structure there is an early active phase which must not be disturbed if memory is to remain intact. Considerable clinical evidence supports the same conclusion. A severe blow on the head leads to a condition known as retrograde amnesia, a loss of memory for the events immediately preceding the accident. Some brain damage, particularly that to the temporal lobe, interferes with recall of recent events but not with remote events. It is argued that these effects are the result of a disruption of an on-going transient process. *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY.

Long-term memory. An anatomical basis for long-term memory exists in the possibility of permanent changes taking place at the synapse. Although there is no conclusive evidence that such changes actually do occur, there is some support for the contention that the number or size of the synaptic knobs which make contact with the cell body increases. This would provide a structural basis for the permanent association between events which is the basis of long-term memory. Regardless of the structural basis of such long-term memories, however, several types of behavioral studies suggest that the process responsible for them differs from that for short-term memories. Those studies mentioned above in connection with short-term memory demonstrated that immediate memory could be disrupted without affecting long-term memory. It has also been demonstrated that memory for recent events survives certain brain operations while memory for more remote events does not. Another rather striking demonstration of this difference is that by electrical stimulation of the brain, which presumably excites structural memory traces, memories for remote events can be revived, though those for immediate events cannot.

At the present stage of knowledge, then, it is reasonable to conclude that two processes constitute the neural basis of memory. Recall of recent events likely depends on a brief, transient neural process, while the recall of more remote events likely depends on a more permanent structural change.

Brain-structure complexity. As the brain has developed from lower to higher species, there has been a progressive increase in the length of time over which the animal is able to remember recent events. In several studies it has been shown that when animals of various species have been tested on similar tests, man can remember longer than monkey, and monkey longer than rat, the length of retention time ranging from several hours in man to a few seconds in rat. Even within the primates, there is a progressive improvement in the length of delay over which material can be retained as the brain increases in complexity from lemur to chimpanzee. One of the changes which occurs in phylogenetic development is an increase in the number and complexity of cortical neurons. This would provide more possibilities for initiating and maintaining the closed-loop reverberating circuits which are thought to be the basis of recent memory. A second development in phylogeny is the proportionately greater increase of association cortex relative to nonassociation cortex. This development, correlated as it is with increased memory span, suggests that memory for recent events may be localized in such cortex. This indeed appears to be the case with respect to prefrontal cortex for it has been shown in many studies that recent memory is impaired following damage to the prefrontal lobes.

These evolutionary developments in the brain do not appear to affect the capacity for long-term memory, however. Conditioned responses, for example, appear to be retained as well by rats as by man. Other material which has been well learned also resists forgetting in all species. Thus, the greater complexity of the primate brain, though providing greater capacity for memory which appears to depend on transient neural processes, does not improve the possibilities of recall of material which likely depends on structural changes occurring in the brain.

Localization of memory. There is considerable disagreement as to the extent to which memory functions are localized and as to which parts of the brain are involved in the functions. Until 1957, evidence for localization of immediate memory in the frontal lobes was convincing. In many studies it had been shown that when monkeys were deprived of their frontal lobes they were unable to retain the memory for a stimulus even for so long as a second. This deficit was not apparent in monkeys with lesions elsewhere in the brain. However, in 1958 investigators have shown a similar impairment from lesions in the head of the caudate nucleus, or in the hippocampus. Furthermore, chimpanzees are not so permanently impaired from such lesions, and man hardly at all. Thus there is some serious question to assigning the function of immediate memory only to the frontal lobes. In fact, in 1956, studies of the effects of bilateral hippocampal lesions in man suggest that the temporal lobe is also involved in memory functions.

Several other sources of data also suggest that a strict localization theory of memory is untenable. It has been observed that if animals are considerably overtrained on a visual discrimination habit, memory for the habit is little affected by a lesion that would severely disrupt a less-well-learned habit. It would appear that the overtraining distributes the structural substrate of memory more widely in the brain. Other investigators have shown that although damage to a certain brain area may result in an animal being unable to recall a specific habit, such damage does not prevent the animal from relearning the habit in fewer trials than it took to learn the problem originally. This suggests, first, that a residual of memory for the habit existed elsewhere, and second that some other area was able to assume these functions. These two

processes, the first technically referred to as equivalence of function, the second as vicarious functioning, argue against strict localization of memory function.

At the same time there is evidence which suggests that there may be some degree of localization. It has been shown that electrical stimulation of the parieto-temporal lobe in conscious patients evokes vivid, complex, remote memories. It does not appear that stimulation of any other area will do this, suggesting that the temporal lobe is the sole store of such memories. Further, memory for language appears to be localized in the dominant hemisphere (left side in right-handed people) in man's brain. Damage to fairly well circumscribed areas in this hemisphere leads to an impairment of language function termed aphasia. Impairment of the sensory aspects of language such as inability to recognize the spoken or written language is a type of aphasia called agnosia, while impairment of the motor aspects of language such as an inability to speak or write the language is called apraxia.

There is considerable debate as to whether or not these disabilities represent memory loss uncomplicated by other factors; nevertheless, the dominating feature in the diagnostic picture of individual subjects is an inability to remember one or several aspects of language. Each of these dysfunctions can often be related to fairly specific loci in the brain, thus supporting the contention that specific localization of memory functions exists. See AGNOSIA; APHASIA; APRAXIA.

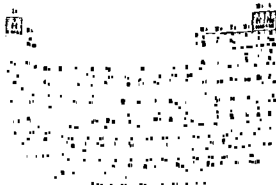
In the light of the present stage of knowledge the most reasonable position to take with respect to localization of memory is that in all probability there are systems within which there is some degree of equipotentiality, such that although a lesion in any one part may have the effect of disrupting immediate recall, residual memory for the event remains, making the restoration of memory possible. On the other hand there is considerable localization of specific memories within these systems. However, the limits of these systems are not yet known.

Conclusion. Although, as observed by Aristotle, it is only through the process of memory, the power of preserving and reproducing presentations and of referring them to past experiences, that creatures are able to advance to generalized knowledge, relatively little is known with any degree of certainty about this basic process. In general, satisfactory methods have been worked out for studying the behavioral correlates of memory, but much remains to be done to clarify essential relationships in this respect. Much less satisfactory are the techniques for studying the neural correlates of memory, and correspondingly less certainty can be placed on the information concerning these aspects of memory. Nevertheless, it is certain that the pace of exploration in this field has been considerably accelerated and that the understanding of memory increases each day. [H.F.RO.]

Bibliography: D. O. Hebb, *A Textbook of Psychology*, 1958; S. S. Stevens (ed.), *Handbook of Experimental Psychology*, 1951; J. Underwood, Interference and forgetting, *Psychol. Rev.*, 64(1): 49–60, 1957; R. S. Woodworth and H. Schlosberg, *Experimental Psychology*, rev. ed., 1954.

Mendelevium

Element number 101, mendelevium, Md, is a member of the actinide series of elements. The element does not occur in nature; it was discovered and is prepared by artificial nuclear transmutation of a lighter element. In 1952, A. Ghiorso, B. G. Harvey, G. R. Choppin, S. G. Thompson, and G. T. Seaborg



bombarded a small amount of Es^{253} with helium ions accelerated in a cyclotron. Einsteinium and helium nuclei combined, and a neutron was emitted to give the mass 256 isotope of element 101. This experiment constituted the discovery of the element which was named in honor of Dimitri Mendeleev, the founder of the periodic table of the elements. Mendelevium-256 decays by capture of an orbital electron with a half-life of about ½ hour. It is quite certain that all isotopes of mendelevium will prove to be radioactive; maximum half-lives will probably be of the order of days. See ACTINIDE ELEMENTS; RADIOACTIVITY.

The amount of mendelevium available for experimentation is much less than one-millionth of a gram and studies of the element's chemical properties have been limited to a tracer scale. Studies of the behavior of mendelevium in ion-exchange chromatography indicate that it exists primarily in the 3+ oxidation state characteristic of the actinide elements. Mendelevium is desorbed from cation-exchange resin at a rate different from that of other actinides when a complexing agent is passed through the column of resin. This provides the basis for the isolation and identification of mendelevium in the presence of other actinide elements. See ION EXCHANGE; NUCLEAR CHEMISTRY; NUCLEAR REACTION; TRANSURANIUM ELEMENTS. [S.C.T.]

Bibliography: J. J. Katz and G. T. Seaborg, *The Chemistry of the Actinide Elements*, 1958.

Mendelism

The basic laws of inheritance discovered by Gregor Mendel, a monk in the Augustinian monastery at Brünn, Moravia. He presented the results of his experiments and his theoretical generalizations on heredity in a paper read before the Natural History

Society of Brünn in 1865 and published his work the following year. The paper was neglected until 1900 when H. DeVries in Holland, C. Correns in Germany, and E. von Tschermak in Austria independently obtained similar results, found Mendel's original work, and brought its fundamental significance to world attention. This was the beginning of the science of genetics. The principles of inheritance which Mendel discovered have since been shown to apply generally among living things that reproduce sexually.

Mendel's laws were formulated to explain the basis for results of his experiments on the inheritance of seven pairs of alternative characteristics in the garden pea. He studied each character separately by making hybrids between plants which were constant for one of a pair of alternative traits such as round versus wrinkled seeds, green versus yellow pods, and tall versus short stems. In the first hybrid generation Mendel noted that each character of the hybrid resembles that of one of the parental forms. These characters he called dominant; the contrasting hidden ones, recessive.

Law of segregation. In a second generation, bred by self-pollinating the hybrid, Mendel observed that both the dominant and the recessive character appeared unaltered and in the proportion of three dominant to one recessive. He extended his experiments to subsequent generations and found that plants showing a recessive characteristic bred true; whereas among the dominants one bred true

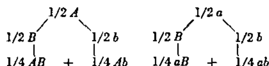
inheritance by units—now called genes—which occur paired in each parent and are separated so that they occur singly in the mature reproductive cells, the gametes. The units become paired once again following the union of gametes in the formation of the next generation. When the paired units are dissimilar a hybrid results. In the hybrid the paired units remain unaltered and subsequently are segregated into approximately equal numbers of gametes. These recombine randomly with reference to the particular units they contain in forming the next generation. See GENE; RECOMBINATION, GENETIC.

Diagrammatically this may be represented as follows

Parent generation:
AA (dominant) aa (recessive)
Gametes of parents:
A a
First hybrid generation:
Aa (resembles dominant parent)
Gametes of hybrid:
$(1/2\,A + 1/2\,a) \times (1/2\,A + 1/2\,a)$
Eggs Pollen
Second hybrid generation:
1

Law (
studyin

only one essential characteristic Mendel investigated the offspring of hybrids involving several alternative characters. To cite one experiment, he pollinated a seed parent of round, yellow peas (AA and BB, both dominant characters) with pollen from a parent plant of wrinkled green peas (aa and bb). The first hybrid generation formed round, yellow seeds and the second hybrid generation yielded seeds of four sorts in the following numbers: 315 round and yellow, 101 wrinkled and yellow, 108 round and green, and 32 wrinkled and green. Mendel adduced that this ratio resulted from independent assortment of the hereditary units controlling these two characters. That is, in the formation of the gametes of the hybrid 1/2 contain A and 1/2 a; similarly 1/2 contain B and 1/2 b. Each pair is sorted out independently of the other so that the hybrid produces four kinds of gametes in approximately equal numbers.



The four kinds of gametes unite at random to form nine different genetic types in the second generation. With dominance, only four differently appearing types are observed. These occur in the following proportions

Genetic types	*Apparent type with dominance*
$1/16\,AABB + 2/16\,AABb + 2/16\,AaBB + 4/16\,AaBb$	$9/16\,A\text{-}B\text{-}$
$1/16\,AAbb + 2/16\,Aabb$	$3/16\,A\text{-}bb$
$1/16\,aaBB + 2/16\,aaBb$	$3/16\,aaB\text{-}$
$1/16\,aabb$	$1/16\,aabb$

Thus Mendel's hypothesis predicts four differently appearing types in the proportions 9 round, yellow seeds:3 wrinkled, yellow seeds:3 round, green seeds:1 wrinkled, green seed. It can be seen that the experimental results cited above fit this expectation.

Mendel devised further tests of his hypotheses by crossing hybrids back to their parents. The results fulfilled expectations and added further proof of the validity of his laws of inheritance. Modern genetics has extended and modified Mendelism, but it remains basically the same as originally presented by Mendel. See GENETICS. [H.H.SM.]

Bibliography: L. C. Dunn (ed.), *Genetics in the 20th Century*, 1951; H. Iltis, *Life of Mendel*, 1932; E. W. Sinnott, L. C. Dunn, and T. Dobzhansky, *Principles of Genetics*, 5th ed., 1958.

Menhaden

A member of the herring family, *Brevoortia tyrannus*, sometimes called mossbunker or porgy. This



Menhaden, *Brevoortia tyrannus*, length to 18 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

every industry uses some product of the menhaden fishery. The menhaden is one of the most important sources of food for many predatory animals in the Atlantic and is consumed by the millions every year. Like other herring, these fishes subsist on plankton.

Each female produces 100,000 or more eggs per season, which float just below the surface. Menhaden are relatively small, averaging under 12 in. in length, but some individuals attain a length of 18 in. *See* CLUPEIFORMES [J.D.B.]

Meninges

The three membranes—the dura mater, arachnoid membrane and pia mater—that cover the brain and spinal cord.

The outermost, the dura mater, is a tough, fibrous double-layered structure that is adherent to the skull. The inner layer of the dura mater sends separating sheets between the cerebral hemispheres and also between cerebrum and cerebellum. It also contains large venous sinuses and forms sheaths for nerves leaving the skull. The middle layer, the arachnoid, is a delicate serous layer loosely investing the brain. Below this is the spongy subarachnoid cavity which contains the circulating cerebrospinal fluid. The innermost layer, or pia mater, is a vascular layer which closely follows each brain convolution. All together, the meninges furnish protection, blood supply, drainage, and cerebrospinal channels for the brain. The meninges around the cord are less complex but basically similar to and continuous with those of the brain. *See* BRAIN; SPINAL CORD. [E.C.ST.]

Meningitis

An inflammation of the delicate membranes, or meninges, which cover the brain and spinal cord. Although principal causes of meningitis are bacterial and viral infections, a yeast (*Torula*) has also been implicated on occasion. Common clinical manifestations include fever, headache, and stiffness of the neck and back. Examination of the spinal fluid yields evidence of inflammation in that there is an increased number of leukocytes and amount of protein. Sometimes the causative microorganism can be found in the spinal fluid. Common causes of bacterial meningitis are meningococcus, ...ococcus, and *Hemophilus influenzae*. Causes ...l meningitis include such agents as lymphocytic choriomeningitis, mumps, poliomyelitis, and the Coxsackie and ECHO groups. A yeast organism of the *Torula* species can cause yeast meningitis. Without specific treatment the fatality rate is high in bacterial meningitis. Such treatment includes specific chemotherapeutic drugs which are effective against particular bacteria. Recovery is usually spontaneous in viral meningitis. *See* ANIMAL VIRUS; BACTERIOLOGY, MEDICAL; MYCOLOGY, MEDICAL. [P.B.BE.]

Meningococcus

A gram-negative diplococcus, the cause of epidemic cerebrospinal meningitis, typically a suppurative disease of the meninges. The bacterium is also known as *Neisseria meningitidis* or *Neisseria intracellularis*.

The meningococcus is an aerobic, gram-negative, nonmotile, nonsporogenous diplococcus about 1 μ in diameter, although it varies in size. Pairs and tetrads, slightly flattened on adjacent sides, or short chains may be observed microscopically. It closely resembles the gonococcus. A capsule is observed microscopically when swollen in the presence of specific antibodies as in the Quellung reaction. In stained films from meningeal exudate, the meningococcus is characteristically seen in the cytoplasm of the polymorphonuclear leukocytes and extracellularly.

The meningococcus is susceptible to the sulfonamides, penicillin, and to the broad-spectrum antibiotics such as the tetracyclines. The latter are capable of crossing the blood-brain barrier in sufficient concentration to inhibit multiplication of the organism. Man is the sole host of the meningococcus. It may cause pharyngitis. Healthy persons or convalescent patients may carry the infection in the naso-pharyngeal mucosa for long periods. *See* BACTERIOLOGY, MEDICAL; GRAM-NEGATIVE DIPLOCOCCI; GRAM'S STAIN; MENINGITIS; QUELLUNG REACTION. [C.M.C.]

Menopause

The natural physiologic cessation of menstruation. Also known as the climacteric, the menopause usually occurs in the last half of the fourth decade of the human female's life but wide variation exists. Decreasing ovarian function and a shift in hormonal balance are responsible. Although atrophic changes occur in the ovaries, uterus, vagina, and other female organs, these are not often immediate or pronounced. Psychic disturbances are common, however, particularly when the menopause is interpreted as the end of femininity or as a mark of sexual inadequacy. In reality, however, there is little direct relationship between these qualities and the naturally declining ability to procreate. Hot flashes, sweating, and various other vasomotor, nervous, and psychic disturbances may occur or may be practically absent. *See* HORMONE.

Sexual intercourse may continue without any perceptible change, and indeed fertilization during or after menopause is not rare since absence of

menses does not necessarily preclude ovulation.

Premature menopause may be seen in younger women, especially after a severe illness or prolonged lactation. Artificial menopause results from surgical removal of the organs or may follow extensive x-ray or other radiation treatment.

Hormone therapy and other medications are effective in most cases where symptoms are troublesome, particularly when psychic factors have been evaluated and alleviated as much as possible.

[E.C.ST.]

Menstruation

Periodic sloughing of the uterine lining in women of reproductive age. A complex hormonal regulation is the principal control mechanism. The developing ovarian follicle, stimulated by follicle-stimulating hormone (FSH) of the pituitary gland, secretes estrogenic hormones in increasing amounts during the first half of the intermenstrual period. The estrogens cause proliferation of the uterine mucosa so that it thickens and shows vessel and gland enlargement. The ovum is extruded from the ovarian follicle during the middle of the cycle, presumably in response to pituitary control, and the remaining portion of the follicle is called the corpus luteum. This begins to secrete progesterone and other hormones which further affect the uterus in preparation for pregnancy. In the absence of fertilization, the corpus luteum deteriorates, progesterone secretion drops, and the greatly thickened and vascularized mucosa is sloughed off. *See* ESTROGEN; HORMONE; HYPOPHYSIS; PROGESTERONE.

Many other factors play a part in this process and wide normal and abnormal variations are encountered. The onset of menstruation, or menarche, occurs usually between the ages of 9–16, with the majority falling in the 12–14 age group. A normal cycle ranges from 21 to over 60 days in different women, with an average of 28 days. Duration of flow is most commonly 3–6 days; amount of flow varies individually and also, within arbitrary limits, from one period to another. Ovulation occurs most often 10–14 days after onset of menstruation in women with a cycle of 24–35 days. Cessation of menses, the menopause, occurs predominantly in women 45–50 years old.

Fertilization normally occurs within a few hours to several days following ovulation, after which time the ovum deteriorates if conception does not occur. *See* FERTILIZATION.

The term menstruation is used to denote the cyclic proliferation and sloughing of the uterine mucosa in humans and a few higher primates. Estrus, in contrast, refers to either the less regular or more widely spaced changes in reproductive organs in lower forms or, more specifically, to the period of intense sexual desire or activity in the females of such lower mammals.

In various species cyclic activity, both functional and sexual, may appear in different patterns. In monestrous animals, one period of sexual activity occurs each year. In polyestrous species, there may be occasional, erratic activity, or it may occur at regular intervals, varying from 2–20 or more times a year. Only higher anthropoids and humans show true menstruation, but bleeding and sloughing of vaginal cells occurs in several other forms, such as the cat and the rabbit. In humans, ovulation occurs between menstrual periods, but in many animals either it does not occur until copulation is achieved, or it may not occur during every cycle. In some forms ovulation is coincident with sexual activity. The duration of the estral cycle for representative species is as follows: chimpanzee, 36 days; macaque monkey, 27 days; sow, 21 days; cow, 20 days; sheep, 16 days; rat and mouse, 4–6 days. In some of these, however, a quiescent, or anestral, period will intervene.

[E.C.ST.]

Mental deficiency

A condition in an individual which is characterized by intellectual retardation, social inadequacy, and persistent dependency. It is evidence of a developmental arrest, and is apparent from birth or an early age. It is of lifetime duration and is resistant to change through training, treatment, or other presently known modifying agents. The causes range from known genetic factors to metabolic dysfunction and specific injury to the central nervous system.

It is estimated that 1–4½% of the general population is mentally deficient. Even the most conservative estimate suggests a population of handicapped far beyond that caused by other debilitating diseases or conditions. The incidence in the public school population has been reported as 25 in 1000. The U.S. Selective Service screening process, from the initiation of the program until the end of World War II, rejected 716,000 individuals aged 18–37 years for military service, for reasons of mental deficiency. In the United States during 1951 public and private institutions for the mentally deficient showed approximately 131,000 residents on their rolls; and during the same year there were about an equal number of retarded children in special public school classes.

Diagnosis and classification. Diagnosis is frequently established in the preschool child by evidence of delayed maturation in the areas of adaptive, social, and verbal behavior. The retarded child of school age is slow to learn or incapable of learning, and tends to develop frustration when he cannot meet the demands of his environment. In adulthood, vocational limitation is evident and the need for protective supervision persists beyond the usual age of social emancipation (*see* FRUSTRATION; LEARNING THEORIES).

Mental deficiency is divided into a number of subgroups on the basis of cause, degree of manifest intellectual retardation, and ability to be trained. From the aspect of cause, the mentally deficient can be classified into two groups: genetic, also called endogenous, or primary, where certain hereditary factors are recognized as having the fun-

damental role; and acquired, called exogenous, or secondary, where etiology is attributed to nonhereditary factors. There are two classes of inherited defect. The first includes persons with generalized mental inferiority, predicated upon multiple genes and resulting in a position of marginal or low status in terms of the normal distribution curve of intelligence. This group is known as familial, reflecting the retardation and social inadequacy of a number of members in a single family. The second class of hereditary deficiency results, as a rule, from a single mutant gene which causes an abnormal neural or metabolic manifestation and, in some cases, a maldevelopment of the skull. Typical of the metabolic conditions are galactosemia, a defect in the metabolism of galactose, a carbohydrate, Tay-Sachs disease, a defect of the lipid metabolism of the brain cells; glycogenosis, a defect of glycogen, or animal starch, metabolism; and gargoylism, involving more than one metabolic defect. Such metabolic disorders actually injure the brain or interfere with its function. Also typical of this group of deficiencies are microcephaly, premature [illegible]

[illegible] LISM). Acquired defects may result from events occurring during the prenatal, natal or paranatal, and postnatal periods. Typical of the prenatal conditions are infection (such as maternal rubella or German measles), the adverse effects of Rh incompatability in the parents, and endocrine disorders like cretinism. Birth injuries and associated adverse events, such as anoxia, are typical of the causative conditions of the paranatal period. Infectious disease, such as encephalitis, and physical accidents in infancy are among the postnatal causes which may produce mental deficiency (*see* ARBOR VIRAL ENCEPHALITIDES; BLOOD GROUPS; CRETINISM; GERMAN MEASLES).

The estimates of distribution of the causes of [illegible]

[illegible] regarding the etiology of mental deficiency is revealed by the large residual percentage of unclassifiable or unknown causes.

Degrees of Intelligence. Mentally deficient individuals are often classified in terms of their mental age, or intelligence quotient (IQ). The highest group, classified by the terms *moron*, mild, and high-grade, falls in general between IQ limits 50–80 and encompasses approximately 82% of the known mental defectives. The next lowest group in degree of retardation, with IQ limits 25–49, is known variously as imbecile, moderate, and middle-grade, and includes approximately 14% of the known mental defectives. The lowest group, with demonstrable IQs generally below 25, is known as idiot, severe, or low-grade, and includes approximately 4% of the known cases.

Classification terms, more commonly used in the public schools and roughly corresponding to those referred to above, are educable, trainable, and nontrainable. The educable are the mentally deficient who, by special training, can be expected to attain comparative vocational and social independence at maturity. The trainable are those who, with special training, are capable of a minor degree of self-help and social participation in a sheltered environment.

Extensive organization of associations for retarded children has resulted in the development of resources for the mentally deficient other than the older programs of residential facilities and public education. Among these resources are diagnostic clinics, sheltered workshops, special recreational programs, and nursery classes for children not yet ready for the public schools. *See* ABNORMAL BEHAVIOR; MONGOLIAN IDIOCY; PHENYLPYRUVIC OLIGOPHRENIA. [M.C.K.]

Bibliography: C. E. Benda, *Mongolism and Cretinism*, 2d ed., 1949; R. L. Masland, The prevention of mental retardation: a survey of research, *Am. J. Mental Deficiency*, 62(6):991–1115, 1958; S. B. Sarason, *Psychological Problems in Mental Deficiency*, 2d ed., 1953; S. B. Sarason and T. Gladwin, Psychological and cultural problems in mental subnormality: a review of research, *Am. J. Mental Deficiency*, 62(6):1115–1305, 1958; R. F. Tredgold and K. Soddy (eds.), *A Textbook of Mental Deficiency*, 9th ed., 1956.

Mental illness

A term used to indicate any form of mental aberration, but usually referring to a chronic or prolonged disorder in which there are wide deviations from the normal.

Few subjects have been as controversial as that of mental illness. This is true of classifications of mental disorders, diagnoses, methods of treatment, and relationships to social and economic systems. Even more disorganized, fragmentary, and fallacious have been the views of the general public on matters of mental illness, much of this being the result of superstition and misunderstanding passed down over years. Some confusion arises because in many mental disorders there are few, if any, organic tissue changes which can be used to form a basis of understanding. Most of these mental aberrations are the result of disorders in behavior and there are many theories concerning behavior of the normal and abnormal individual. *See* ABNORMAL BEHAVIOR.

Although systems of classification differ, the following categories may be recognized.

Three major groups of mental disorder are (1) those caused by, or associated with, impairment of brain tissue function; (2) those caused by mental deficiency of hereditary, familial, or unknown cause; and (3) those of psychogenic origin or in which there is no physical or structural alteration in the brain.

Conditions associated with impaired brain tissue function may be acute or chronic. Acute types may follow trauma, infections, poisonings, vascular disorders, metabolic or nutritional disturbances,

convulsive states, and other induced damage to the brain, including that produced by the growth of tumors.

Chronic conditions that show physical changes in brain tissue are caused by much the same kind of agent as those which produce acute cases; the main difference is one of degree and length of time required to produce brain damage.

Mental deficiency commonly results from adverse hereditary or familial traits, but may be found in many cases in which no specific causes or contributing factors are apparent. *See* HUMAN GENETICS.

Psychogenic brain disorders can be grouped into several main categories. The psychoses are marked by disorganization of the personality so that contact with reality, interpersonal relationships, and the capacity for productivity are altered to the extent that impairment of one or more of these functions is characteristic. *See* PSYCHOSIS.

Involutional types, such as schizophrenia and paranoid reactions, and affective states, such as manic depression, are the principal forms of psychosis. *See* MANIC DEPRESSIVE PSYCHOSIS; PARANOID STATE; SCHIZOPHRENIA.

The psychoneuroses are more common, benign conditions in which a person fails to adapt successfully to some psychic situation. This produces anxiety, considered the hallmark of the psychoneurotic; yet the anxiety is itself a symptom of a more deep-seated struggle. The type of defense displayed by the patient to a stressful situation may be single or mixed. Anxiety reactions, depressive reactions, obsessive-compulsive, phobic, conversion, and dissociative mechanisms are commonly seen, but seldom in a completely pure form. *See* HYSTERIA; OBSESSIVE COMPULSIVE REACTION; PHOBIC REACTION; SOMATIZATION.

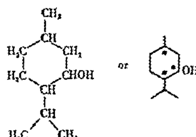
Other common and variably severe mental disorders are those in which there is a defect or abnormal trend in the personality of the individual. These include, among other conditions, those of the inadequate personality, the emotionally unstable, and the various types of addiction, sexual deviation, and other antisocial features.

A group of disorders sometimes called the psychophysiologic states are marked by the excessive or abnormal reactivity of some system or tissues of the body under control of the autonomic nervous system. These conditions, affecting the skin, gastrointestinal tract, and other organs, comprise the bulk of what has come to be called the psychosomatic diseases. The organ responses are believed by many to represent a form of defense against anxiety in most cases. Certain instances of hypertension, stomach ulcers, dermatoses, and genitourinary disturbances are common examples.

Treatment of mental illness is almost an individual matter. The use of shock therapy, drugs, psychotherapy, and other types of treatment is widespread, but the demand for qualified psychiatrists exceeds the supply. See PSYCHOPHARMACOLOGIC DRUGS. [E.G.ST.]

Menthol

A monocyclic, saturated, secondary terpene alcohol. Menthol contains three (starred) asymmetric carbon atoms. It can exist in four externally com-



Menthol

pensated and eight optically active forms. The only forms encountered in nature are the *l*-menthol and *d*-neomenthol. By far the most important is *l*-menthol, found as the main constituent (50–65%) in peppermint oil.

Commercial *l*-menthol is isolated principally from the oil of *Mentha arvensis* grown in Japan. The process involves cooling the oil and purifying the crystals formed. It possesses a distinct peppermint flavor and gives the impression of cooling the mouth and skin. Appreciable amounts of the racemic mixture of the optical isomers of menthol are synthesized commercially.

l-Menthol is widely used as a flavoring ingredient in tooth pastes, mouth washes, and cigarettes, and as a rubefacient, and cooling agent in chest rubs. *See* TERPENE. [E.L.S.]

Mercaptan

One of a group of organosulfur compounds which are also called thiols or thio alcohols and which have the general structure, RSH. Aromatic thiols are called thio phenols, and biochemists often refer to thiols as sulfhydryl compounds. The unpleasant odor of volatile thiols classes them as stenches, but the odors of many solid thiols are not unpleasant.

Mercaptans (1) form salts with bases, (2) are easily oxidized to disulfides and higher oxidation products such as sulfonic acids, (3) react with chlorine (or bromine) to form sulfenyl chlorides (or bromides), and (4) undergo additions to unsaturated compounds, such as olefins, acetylenes, aldehydes, and ketones. The insoluble mercury salts (mercaptides) are used to isolate and identify mercaptans.

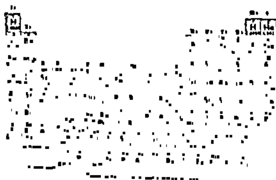
Some important amino acids, such as cysteine and methionine, and many pharmaceutical and industrial products contain the mercapto group. Thus, thiosalicylic acid is used in synthesizing the germicide, Merthiolate; 2-mercaptobenzothiazole is a rubber vulcanization accelerator; 2,3-dimercaptopropanol (British anti-Lewisite) is an antidote for arsenic poisoning; and 6-mercaptopurine is of interest in cancer chemotherapy. The occurrence and removal of mercaptans from petrole

is industrially important. Even the odor of mercaptans finds use; traces of mercaptans added to dangerous gases act as warning agents in case of leaks.

The name thio alcohols suggests that mercaptans are similar to alcohols. Although some properties are analogous, there are decided differences, related to (1) the greater acidities of thiols, (2) the ease with which mercaptans are oxidized, and (3) the ability of mercaptans to enter free-radical reactions. Difference (2) probably accounts for the absence of thiols, as such, in contrast to alcohols, in nature. *See* ORGANOSULFUR COMPOUND; PETROLEUM PROCESSING; RUBBER; SULFENYL CHLORIDES. [N.K.]

Mercury (element)

Chemical element number 80, mercury, Hg, has been known since ancient times. Mercury is a silver-white liquid at room temperature (melting point, −38.87°C); it boils at 356.9°C under atmospheric pressure. It is a noble metal that is soluble only in oxidizing solutions. Solid mercury is as soft as lead. The metal and its compounds are



very toxic. With some metals (gold, silver, platinum, uranium, copper, lead, sodium, and potassium, for example) mercury forms solutions called amalgams. With other elements, such as iron, cobalt, nickel, manganese, antimony, and silicon, it does not react. *See* AMALGAM.

In its compounds, mercury is found in the 2+ or 1+ oxidation states, for example, $HgCl_2$ or Hg_2Cl_2. In both cases, the mercury atoms are doubly covalently bonded, for example, Cl—Hg—Cl or Cl—Hg—Hg—Cl. Some mercury(II) salts, for example, $Hg(NO_3)_2$ or $Hg(ClO_4)_2$, are quite soluble in water, and dissociate normally. The aqueous solutions of these salts react as strong acids on account of hydrolysis. Other mercury(II) salts, for example, $HgCl_2$ or $Hg(CN)_2$, also dissolve in water, but exist in solution as only slightly dissociated molecules. This is related to the tendency of mer- [illegible] —Hg—NH—CO—CH_3. In complex compounds, for example, $K_2(HgI_4)$, mercury often has three or four bonds.

Uses. Metallic mercury is used as a liquid contact material for electrical switches, in vacuum technology as the working fluid of diffusion pumps, for the manufacture of mercury-vapor rectifiers, thermometers, barometers, tachometers, and thermostats, and for the manufacture of mercury-vapor lamps. Mercury-vapor lamps serve as sources for ultraviolet light because they emit a line spectrum with the following principal lines: 365.0/366.3 mμ, 404.7 mμ, 435.8 mμ, 546.1 mμ, and 577.0/579.0 mμ. For micro gas analyses, mercury is most often used as a sealing liquid for the evolved gases. Very large amounts of mercury are used as the electrode material for the electrolysis of aqueous solutions of alkali halides for the manufacture of chlorine and sodium hydroxide. Also, it finds application for the manufacture of silver amalgams for tooth fillings in dentistry. *See* MERCURY-VAPOR LAMP; MERCURY-VAPOR RECTIFIER.

Some mercury salts serve as catalysts for organic chemical reactions. Fulminate of mercury, $Hg(CNO)_2$, is used as a primer for explosives. Some complex salts, $Cu_2(HgI_4)$ and $Ag_2(HgI_4)$, find application as temperature indicators, for example, for detecting overheated machine parts since they change color on heating. A large number of the mercury compounds have been used for hundreds [illegible] cation as a disinfectant in dilutions of 1:1000 in water. The so-called yellow mercury salve contains 5% HgO and is used to treat conjunctivitis. Ammoniated mercury ointment contains 5–10% $HgNH_2Cl$ and is used in the treatment of various skin diseases.

Of importance in electrochemistry are the standard calomel electrode, used as the reference electrode for the measurement of potentials and for potentiometric titrations, and the Weston standard cell, with which the accurate measurement of potential sources is possible. In the calomel electrode and in the Weston standard cell, metallic mercury is used in contact with solutions which contain mercury(I) salts as the solid phase. *See* CALOMEL ELECTRODE.

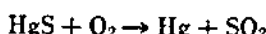
Natural occurrence. Mercury is commonly found as the sulfide, HgS; frequently as the red cinnabar, less often as the black metacinnabar. Important deposits are found in Spain, Italy, California, Nevada, Oregon, Texas, Mexico, Canada, Brazil, Peru, China, Japan, Russia, Hungary, Yugoslavia, and Germany. A less common ore is the mercury(I) chloride found in Texas. Occasionally, the mercury ore contains small drops of metallic mercury. *See* CINNABAR.

The reserve of mercury at Almadén, Spain, has been estimated at 40,000 tons, that in Idria, Italy, at about 20,000 tons. The ore found in Spain has the highest mercury content, with an average of 0.5–1.2% Hg, sometimes as much as 10% Hg.

It has been estimated that the outer 16-km thick layer of the earth's crust contains 2.7×10^{-6}% Hg. Mercury is less abundant than platinum, uranium,

tantalum, cesium, silver, osmium, palladium, and indium. Nevertheless, it is not regarded as a rare element because it is found in highly concentrated deposits, and hence is readily available.

Metallurgical extraction. The separation of mercury metal from ores is accomplished most often by heating in an air stream in a rotary kiln or shaft furnace. In this process, the sulfide ore is roasted



$$HgS + O_2 \rightarrow Hg + SO_2$$

or reduced to metallic mercury by the addition of iron



$$HgS + Fe \rightarrow Hg + FeS$$

or of quicklime



$$4HgS + 4CaO \rightarrow 4Hg + 3CaS + CaSO_4$$

The metal vapor is carried over with the combustion gases and condenses in vertical clay pipes having open bottoms and standing in water. The mercury metal collects under the water in a layer of condensation products called mercurial soot, which also contains mercury salts, soot, and tar, and from which the metal is separated by filtration through a cloth. The mercurial soot is pressed together with quicklime, whereby additional liquid mercury is obtained. The residue from the compression is roasted together with new mercury ore. The final purification of the mercury metal is best done by vacuum distillation, by dropping it into a 1.5-m layer of 5% nitric acid or by pressing it through leather.

The mercury metal is sold in iron flasks 45 cm long and 10 cm thick. Each flask contains 76 lb (34.5 kg) of metal (34.5 kg is the old Spanish hundredweight).

The world production of metallic mercury is about 5000 tons per year in peacetime and rose to about 9000 tons per year during World War II in 1941 (Fig. 1). Three-fourths of the world peacetime production is from Spain and Italy together, in about equal parts. In 1954, Italy produced 2070 tons of mercury, Spain 1485 tons, United States 656 tons, Yugoslavia 551 tons, Japan 466 tons, and Mexico 354 tons. The price of 1 kg of metallic mercury is about $10.

Physical and chemical properties. The element mercury has atomic number 80, and atomic weight

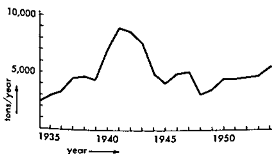


Fig. 1. Total mercury production.

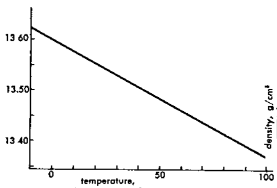


Fig 2. Variation of mercury density with temperature.

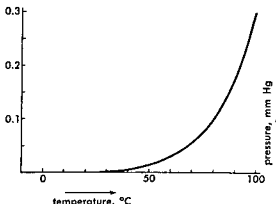


Fig 3 Vapor pressure-temperature relationship of mercury.

200.61. It occurs naturally as a mixture of the following isotopes in the proportions shown: 202 (29.8%), 200 (23.1%), 199 (16.9%), 201 (13 2%), 198 (10.0%), 204 (6.8%), and 196 (0.15%).

The thermal expansion of mercury metal is relatively large, and between 0°C and 100°C the thermal expansion approximates that of a gas. The specific gravity of metallic mercury is 13.54616 at 20°C. The variation of specific gravity with temperature is shown in Fig. 2.

The vapor pressure of mercury rapidly increases with the temperature (Fig. 3). In 1 m³ of air there are accordingly 14 mg of Hg at 20°C and 2.42 g at 100°C, if the air is saturated with mercury vapor.

The thermal conductivity of mercury at 0°C is only 2.2% of that of silver; the electrical conductivity at 0°C is only 1.58% of that of silver. By international definition, a column of mercury 1 mm² in cross-sectional area and 106.3 cm long at 0°C has an electrical resistance of 1 ohm. *See* OHM.

The surface tension of liquid mercury is 505 dyne/cm, six times greater than that of water in contact with air. Hence, mercury does not wet surfaces with which it is in contact.

Mercury metal and mercury compounds are all diamagnetic.

is industrially important. Even the odor of mercaptans finds use: traces of mercaptans added to dangerous gases act as warning agents in case of leaks.

The name thio alcohols suggests that mercaptans are similar to alcohols. Although some properties are analogous, there are decided differences, related to (1) the greater acidities of thiols, (2) the ease with which mercaptans are oxidized, and (3) the ability of mercaptans to enter free-radical reactions. Difference (2) probably accounts for the absence of thiols, as such, in contrast to alcohols, in nature. *See* ORGANOSULFUR COMPOUND; PETROLEUM PROCESSING; RUBBER; SULFENYL CHLORIDES. [N.K.]

Mercury (element)

Chemical element number 80, mercury, Hg, has been known since ancient times. Mercury is a silver-white liquid at room temperature (melting point, −38.87°C), it boils at 356.9°C under atmospheric pressure. It is a noble metal that is soluble only in oxidizing solutions. Solid mercury is as soft as lead. The metal and its compounds are very toxic. With some metals (gold, silver, platinum, uranium, copper, lead, sodium, and potassium, for example) mercury forms solutions called amalgams. With other elements, such as iron, cobalt, nickel, manganese, antimony, and silicon, it does not react. *See* AMALGAM.

In its compounds, mercury is found in the 2+ or 1+ oxidation states, for example, $HgCl_2$ or Hg_2Cl_2. In both cases, the mercury atoms are doubly covalently bonded, for example, Cl—Hg—Cl or Cl—Hg—Hg—Cl. Some ... ar ... in ... so ... react as strong acid ... slightly dissociated molecules. This is related to the tendency of mercury to form covalent compounds. There are compounds in which mercury atoms are bound directly to carbon or nitrogen atoms, for example, H_3C—Hg—CH_3 or H_3C—CO—NH—Hg—NH—CO—CH_3. In complex compounds, for example, $K_2(HgI_4)$, mercury often has three or four bonds.

Uses. Metallic mercury is used as a liquid contact material for electrical switches, in vacuum technology as the working fluid of diffusion pumps, for the manufacture of mercury-vapor rectifiers, thermometers, barometers, tachometers, and thermostats, and for the manufacture of mercury-vapor lamps. Mercury-vapor lamps serve as sources for ultraviolet light because they emit a line spectrum with the following principal lines: 365.0/366.3 mμ, 404.7 mμ, 435.8 mμ, 546.1 mμ, and 577.0/579.0 mμ. For micro gas analyses, mercury is most often used as a sealing liquid for the evolved gases. Very large amounts of mercury are used as the electrode material for the electrolysis of aqueous solutions of alkali halides for the manufacture of chlorine and sodium hydroxide. Also, it finds application for the manufacture of silver amalgams for tooth fillings in dentistry. *See* MERCURY-VAPOR LAMP; MERCURY-VAPOR RECTIFIER.

Some mercury salts serve as catalysts for organic chemical reactions. Fulminate of mercury, $Hg(CNO)_2$, is used as a primer for explosives. Some complex salts, $Cu_2(HgI_4)$ and $Ag_2(HgI_4)$, find application as temperature indicators, for example, for detecting overheated machine parts since they change color on heating. A large number of the mercury compounds have been used for hundreds of years as medicines. The compounds of mercury with some organic substances are powerful diuretic substances. $HgCl_2$ or $HgO{\cdot}Hg(CN)_2$ finds application as a disinfectant in dilutions of 1:1000 in water. The so-called yellow mercury salve contains 5% HgO and is used to treat conjunctivitis. Ammoniated mercury ointment contains 5–10% $HgNH_2Cl$ and is used in the treatment of various skin diseases.

Of importance in electrochemistry are the standard calomel electrode, used as the reference electrode for the measurement of potentials and for potentiometric titrations, and the Weston standard cell, with which the accurate measurement of potential sources is possible. In the calomel electrode and in the Weston standard cell, metallic mercury is used in contact with solutions which contain mercury(I) salts as the solid phase. *See* CALOMEL ELECTRODE.

Natural occurrence. Mercury is commonly found as the sulfide, HgS; frequently as the red cinnabar, less often as the black metacinnabar. Important deposits are found in Spain, Italy, California, Nevada, Oregon, Texas, Mexico, Canada, Brazil, Peru, China, Japan, Russia, Hungary, Yugoslavia, and Germany. A less common ore is the mercury(I) chloride found in Texas. Occasionally, the mercury ore contains small drops of metallic mercury. *See* CINNABAR.

The reserve of mercury at Almadén, Spain, has been estimated at 40,000 tons, that in Idria, Italy, at about 20,000 tons. The ore found in Spain has the highest mercury content, with an average of 0.5–1.2% Hg, sometimes as much as 10% Hg.

It has been estimated that the outer 16-km thick layer of the earth's crust contains 2.7×10^{-6}% Hg. Mercury is less abundant than platinum, uranium,

Compound and formula	*Properties and uses*
Mercury(II) nitrate, $Hg(NO_3)_2 \cdot H_2O$	Mp 79°C with decomposition; colorless crystals, or if slightly hydrolyzed by water, yellowish powder; soluble in nitric acid; insoluble in alcohol
Mercury(II) oleate	Yellowish to red liquid; semisolid or solid mass of indefinite composition; slightly soluble in alcohol and ether; insoluble in water
Mercury(II) oxide, HgO	Decomposes at 500°C into mercury and oxygen; soluble in acids; very slightly soluble in water; insoluble in alcohol and ether; exists in a finely divided yellow form which is more reactive than the crystallized red form; both forms have the same crystal structure
Mercury(II) oxycyanide, $HgO \cdot Hg(CN)_2$	Explodes on heating; white crystalline powder; slightly soluble in water
Mercury(II) perchlorate, $Hg(ClO_4)_2 \cdot 6H_2O$	Colorless crystals; hygroscopic; soluble in perchloric acid solutions
Mercury(II) phosphate, $Hg_3(PO_4)_2$	Heavy white or yellowish powder; soluble in acids; insoluble in water and alcohol
Mercury(II) potassium cyanide, $K_2[Hg(CN)_4]$	Colorless crystals, very poisonous; soluble in water and alcohol
Mercury(II) potassium iodide, $K_2(HgI_4)$	Ne[illegible] -luble in water; crystallizes water-free in [illegible] -ules of water, also
Mercury(II) salicylate, $Hg(C_6H_4O)COO$	W[illegible] 59.5% [illegible]
Mercury(II) silver iodide, $Ag_2(HgI_4)$	Yellow pow[illegible] diluted acids, soluble in [illegible] nide and potassium iodide
Mercury(II) subsulfate, $2HgO \cdot HgSO_4$	Heavy lemon-yellow powder, turns red on heating, but yellow again on cooling, soluble in acids, insoluble in water
Mercury(II) sulfate, $HgSO_4$	White crystalline powder, hydrolyzes with water; insoluble in alcohol; soluble in acids
Mercury(II) sulfide, black, HgS	Black powder; insoluble in water, acids, and organic solvents; soluble in sodium sulfide solution; sublimes at 446°C, yielding the red form of HgS, which possesses another crystal structure; the black sulfide is formed by passing hydrogen sulfide gas into mercury salt solutions or by reaction between metallic mercury and sulfur
Mercury(II) sulfide, red, HgS	Cinnabar, red crystals; insoluble in water, acids, and organic solvents, sublimes at 446°C, the red form of HgS is the main mercury ore
Mercury(II) thiocyanate, $Hg(SCN)_2$	White powder, only slightly soluble in water, soluble in alcohol; decomposes on heating
Mercurochrome, $C_{20}H_8O_6Br_2Na_2HgOH \cdot 3H_2O$	Disodium 2,7-dibromo-4-hydroxymercurifluorescein; green scales; soluble in water; insoluble in alcohol, ether, and chloroform; used in 2% aqueous solution as a general antiseptic
Mercury(I) acetate, $Hg_2(CH_3COO)_2$	Colorless plates, decomposes by light or in boiling water into Hg and $Hg(CH_3COO)_2$, slightly soluble in water; insoluble in alcohol and ether
Mercury(I) bromide, Hg_2Br_2	White powder or tetragonal crystals; becomes yellow on heating and white again on cooling; sublimes at 340–350°C; mp 405°C
Mercury(I) chlorate, $Hg_2(ClO_3)_2$	White crystals; explodes with combustible substances, mp 250°C (decomposition); soluble in water and alcohol
Mercury(I) chloride, Hg_2Cl_2	Calomel; white rhombic crystals; insoluble in water, alcohol, and ether, mp 303°C, bp 384°C; becomes black with ammonia ($Hg_2Cl_2 + 2NH_3 \rightarrow Hg + HgNH_2Cl + NH_4Cl$) by yielding finely divided metallic mercury
Mercury(I) chromate, Hg_2CrO_4	Brick-red powder decomposes on heating to yield Cr_2O_3; insoluble in water and alcohol, soluble in concentrated nitric acid
Mercury(I) iodide, Hg_2I_2	Yellow amorphous powder; becomes greenish in light ($Hg_2I_2 \rightarrow Hg + HgI_2$); orange on heating, but yellow again on cooling, insoluble in water, alcohol, and ether; sublimes at 110–120°C; mp 290°C (partial decomposition); bp 310°C
Mercury(I) nitrate, $Hg_2(NO_3)_2 \cdot 2H_2O$	Short prismatic crystals; soluble in small quantities of warm water, in nitric acid, in boiling carbon disulfide, and in methylamine, and slightly soluble in benzonitrile; hydrolyzes with large quantities of water; mp 70°C with decomposition
Mercury(I) sulfate, Hg_2SO_4	White crystalline powder; almost insoluble in water; soluble in hot sulfuric acid and in dilute nitric acid; decomposes on heating
Mercury dimethyl, $Hg(CH_3)_2$	Colorless liquid; very strong poison; bp 92°C; density at 19°C 3.084 g/cm³

Compound and formula	*Properties and uses*
Mercury diphenyl, $Hg(C_6H_5)_2$	Colorless crystals, mp 122°C; decomposes on heating into mercury and diphenyl
Mercury fulminate, $Hg(CNO)_2$	Dark brown crystalline powder; explodes when dry under the slightest shock and by heating; is made by the action of concentrated nitric acid and alcohol on mercury metal
Mercury methyl chloride, CH_3HgCl	Colorless plates; mp 170°C; very strong poison
Millon's base, $(Hg_2N)OH \cdot 2H_2O$	Yellow powder, insoluble in water and organic solvents; exchanges OH^- ions with halide ions, when in contact with aqueous alkali halide solutions

See TRANSITION ELEMENTS. [K FR.]

Bibliography: R. Abegg, *Handbuch der anorganischen Chemie*, vol. 2, 1922; J. Chatt, The addition compounds of olefine with mercuric salts, *Chem. Revs.*, 48(1):7-43, 1951; L. Gmelin and K. Kraut, *Handbuch der anorganischen Chemie*, vol. 5, 1914; E. Krause and A. von Grosse, *Die Chemie der metall-organischen Verbindungen*, 1937; E. G. Rochow, D. T. Hurd, and R. N. Lewis, *The Chemistry of Organometallic Compounds*, 1957; N. V. Sidgwick, *The Chemical Elements and their Compounds*, 1950; F. C. Whitmore, *Organic Compounds of Mercury*, 1921.

Mercury (planet)

The planet nearest to the Sun. It is visible to the naked eye near its greatest angular distances from the Sun (ranging from 18 to 28°) shortly after sunset or before sunrise. Its orbit has a semimajor axis (mean distance to Sun) of 36.2×10^6 miles. Its eccentricity of 0.206, the highest value of all the main planets except Pluto, causes the distance to the Sun to vary from 28.7×10^6 mi at perihelion to 43.6×10^6 mi at aphelion. The sidereal period of revolution is 87.969 days; the mean orbital velocity, 29.9 mi/sec; and the inclination of the orbital plane to the ecliptic is 7.0°, the largest of the main planets except Pluto. *See* PLANET.

The apparent diameter of its disk varies from 5″ at superior conjunction to about 13″ at inferior conjunction. The linear diameter, about 3000 mi, is not very accurately known because of the smallness of the disk and the difficulty of observation. Polar flattening is negligible. The mass, about 0.053 (Earth = 1), is uncertain since, for want of a satellite, the value can be derived only from the perturbations caused by Mercury in the motions of other planets. The mean density, about 5.3 g/cm³, one of the highest among the planets, is even more uncertain since it is the ratio of two quantities (mass and volume), each of which is poorly known. The value of the acceleration of gravity on the surface, about 3.6 m/sec², is also only approximate.

Appearance and rotation. Through a telescope, Mercury presents an appearance generally similar to that of the Moon seen with the naked eye. As an interior planet it also presents phases similar to those of the Moon. Its surface markings, first mapped by G. V. Schi...

bital eccentricity), indicating that the period of rotation is accurately equal to the period of revolution (88 days); this equality was brought about by tidal friction (*see* TIDE). Thus Mercury always presents the same face to the Sun. As a result, the illuminated side of the planet reaches a surface temperature of about 340°C, according to radiometric measurements, while the dark side must be at a temperature of only a few degrees above absolute zero.

Atmosphere. The small acceleration of gravity, high temperature of the hot side, and low temperature of the cold side combine to preclude the formation and persistence of an atmosphere. Molecules of all gases, except perhaps the heaviest, such as argon, must soon exceed the velocity of escape, about 4 km/sec, on the hot side; and any gas, except perhaps the most refractory, moving to the cold side must soon freeze out and condense on the surface of the planet.

Doubtful changes in the visibility of some dark spots have nevertheless been construed as evidence for the presence of a thin atmosphere, dense enough to support occasional faint veils of dust raised from the surface by convection caused by intense solar heating. Measurements of polarized light have also been interpreted to support this inference and to estimate the (maximum) mass per unit area of a possible argon atmosphere. The tenuity of the atmosphere, if any, is also confirmed by the low albedo, about 0.06, similar to that of



Fig. 1. Telescopic aspect of Mercury.

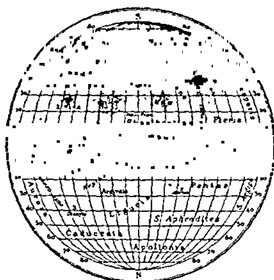


Fig. 2. Map of Mercury, by E. M. Antoniadi.

the Moon, by the absence of any additional absorption lines or bands in the spectrum of sunlight reflected by the planet, and finally by the lack of a bright edge around the disk of the planet when it was photographed projected against the solar corona, only a few minutes of arc from the Sun's limb, on May 11, 1937.

Transits. Because of the inclination of its orbit, Mercury is usually several degrees north or south of the Sun at inferior conjunction; it is only at rare intervals, when the planet is simultaneously at inferior conjunction and near one of the orbital nodes, that a transit in front of the Sun's disk can take place. The Earth passes the line of nodes on May 7 and November 9 so that transits can occur only near these dates. During the May transits Mercury is near its aphelion and therefore nearer to the Earth than during the November transits; consequently the tolerance on the planet's position with respect to the line of nodes for a transit to take place is less stringent in November. As a result the November transits are about twice as frequent as the May transits. The dates of the transits of Mercury between 1951 and 2000 are as follows: 1953, Nov. 14; 1957, May 5; 1960, Nov. 7; 1970, May 8; 1973, Nov. 9; 1986, Nov. 12; 1993, Nov. 5; and 1999, Nov. 15. During some transits of Mercury a bright spot was observed on the dark disk of the planet, and in the nineteenth century, there was some wild speculation (for example, about volcanoes) as to the cause of the phenomenon which is now known to be due to a diffraction effect resulting from excessive diaphragming of the telescopic objective. *See* TRANSIT (ASTRONOMY). [C.D.V.]

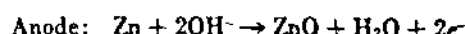
Bibliography: E. M. Antoniadi, *La planète Mercure et la rotation des satéllites*, 1937; P. Moore, *J. Brit. Interplanet. Soc.*, 13, 1954.

Mercury battery

A primary dry-cell battery consisting of a zinc anode, a cathode of mercuric oxide (HgO) mixed with graphite, and an electrolyte of potassium hydroxide (KOH) saturated with zinc oxide (ZnO). With carefully purified materials and balanced amounts of ZnO and HgO, the cell has very low self discharge and makes efficient use of the active materials.

Within the steel can, the active materials are separated by a porous material which prevents migration of conducting particles from the mercuric oxide pellet. Dense dialysis paper and porous polyvinyl chloride have been used for this purpose. The electrolyte is completely absorbed in the active materials, separator, and absorbent materials. The steel can serves as the contact to the HgO. A metal top with a concentric neoprene grommet closes off the top of the can and serves as the contact to the zinc.

The electrochemical system may be written



$$\text{Anode:}\quad Zn + 2OH^- \rightarrow ZnO + H_2O + 2e^-$$

The half-cell potential for a zinc anode in a solution of KOH having a concentration of 7.7 normal (N) has been calculated to be 1.317 volt.



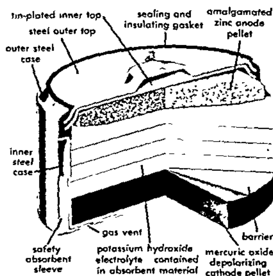
$$\text{Cathode:}\quad HgO + H_2O + 2e^- \rightarrow Hg + 2OH^-$$

The half-cell potential has been calculated to be 0.028 volt. The cell potential, therefore, should be 1.345, which agrees very well with measurement. The over-all cell reaction would be



$$Zn + HgO \rightarrow ZnO + Hg$$

This does not involve the electrolyte. The cell potential, therefore, does not change appreciably with different concentrations of alkali.



Cutaway view of a mercury cell.

When current flows, the ZnO formed in the cell reaction quickly saturates the small amount of electrolyte and then precipitates out. This maintains a constant composition of the electrolyte. Offsetting this is the transport of water away from the anode by the solvated potassium ions. The equilibrium under steady current flow is the result of rather complex exchanges through the separator, which makes the cell-voltage characteristics under load dependent on the initial electrolyte composition.

The mercury cell has a theoretical output of 0.247 ampere-hour per gram (amp-hr/g) of HgO. In practice, the cathode pellet contains about 95% HgO and 5% graphite, having a theoretical output of 0.234 amp-hr/g. The anode is 90% zinc and 10% mercury. This has a theoretical output of 0.738 amp-hr/g.

As built, the cells have slight excess of cathodic capacity. A discharged cell will then have no zinc left to react with the electrolyte and evolve hydrogen. Thus a cell with 12.5 g of cathodic material has 3.6 g of zinc amalgam compared with 3.98 needed for exact balance.

The electrolyte used is about 1 ml/amp-hr. One composition is 100 g KOH, 100 ml H_2O, 16 g ZnO. The actual cell capacity is only slightly less than the theoretical. The over-all cell output is approximately 0.36 amp-hr/lb and 5 amp-hr/$in.^3$

Typical characteristics of a cell on continuous discharge to a final voltage of 0.9 volt at 70°F are as follows:

Hours to 0.9 volt	amp	amp-hr	amp-hr/lb	relative output
10	0.127	1.27	17.4	77%
100	0.019	1.9	26.0	100%

The 10-hour output per pound is about 3 times [illegible] the Leclanche cell. Typical density for the mercury cell is about 3.9 g/ml. For Leclanche cells, the density is about 2. Therefore, the advantage of the mercury cell over the Leclanche cell is about *twice as great* on an equal-volume comparison as on an equal-weight comparison.

Mercury cells, at normal drain rates, have about the same output over a temperature range of 70–140°F. Below 70°F, the output falls rapidly. The output drops 43% at 40°F, 90% at 20°F, and 100% at 0°F.

Heavy drain output is greatly improved at high temperatures. One cell has been reported to have delivered 2.5 amp for over 10 min at 200°F, whereas at 70°F the voltage failed in a few seconds.

[S.EI.]

Bibliography: M. Friedman and C. E. McCauley, The Ruben cell—a new alkaline primary dry cell battery, *Trans. Electrochem. Soc.*, 92:195, 1947; S. Ruben, Balanced alkaline dry cells, *Trans. Electrochem. Soc.*, 92:183, 1947; G. W. Vinal, *Primary Batteries*, 1950.

Mercury-vapor lamp

A vapor lamp in which the discharge takes place in mercury vapor. It is widely used as a source of visible and ultraviolet radiation.

Construction. The mercury-vapor lamp contains an arc tube filled with argon gas and a small amount of pure mercury. The arc tube is usually mounted within an outer bulb of glass. The arc tube itself is usually made of fused quartz. One electrode extends into the tube from one end; a similar main electrode and a smaller starting electrode are at the other end of the tube. The starting electrode is electrically connected through a high resistance to the main electrode at the opposite end of the tube. The outer bulb serves to support the arc tube and provide it with thermal insulation, to prevent atmospheric attack on internal parts, and in some cases, to filter out harmful ultraviolet energy (see Fig. 1).

Starting and operation. The mercury lamp is connected through its socket to the output leads of its ballast, which supplies proper voltage for starting and limits current during operation (*see* VAPOR LAMP). When the ballast circuit is first energized, no current flows, and full starting voltage appears between the starting electrode and the adjacent main electrode. This voltage draws electrons across the relatively short gap, ionizing some of the argon gas in the tube and setting up a glow discharge between these two electrodes. The resistor in the circuit limits current to a few milliamperes. The ionized argon gradually diffuses through the tube, reducing the resistance in the

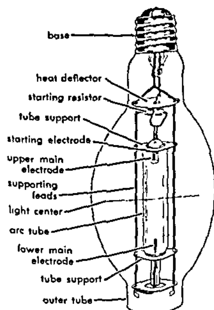


Fig. 1. Mercury-vapor lamp.

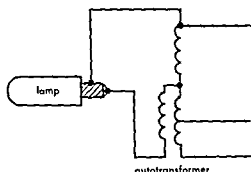


Fig. 2. Wiring diagram for mercury-lamp ballast.

gap between main electrodes. When resistance is low enough, an arc strikes across the main electrodes. The heat from the arc vaporizes the droplets of mercury, and they become ionized current-carriers as electrons in the arc bombard the vaporized mercury atoms. When all the mercury is vaporized, the current in the arc may reach several amperes. With this current flowing in the ballast, it no longer produces sufficient voltage to maintain the initial glow, which is extinguished. The arc is then maintained across the main electrodes with its current limited by the ballast (see Fig. 2).

If the arc is extinguished by a momentary power failure or deliberate disconnection of the power supply, it cannot be restarted immediately. While the arc tube is still hot, the pressure created by the still-vaporized mercury is too high to permit the formation of the glow discharge at the starting electrode. A cooling period is necessary to allow the mercury to condense on the arc-tube walls, lowering pressure sufficiently for the process to begin again.

Radiation. Radiation from the mercury arc is confined to four specific wavelengths in the visible portion of the spectrum and several lines in the ultraviolet region. Mercury vapor pressure—governed by arc-tube volume, quantity of mercury and current—can be regulated to concentrate as much radiation as possible in the wavelengths most desired for the intended application of the lamp.

The visible radiation from mercury lamps is all concentrated in the blue, green, and yellow-green regions of the spectrum. This gives light from ordinary mercury lamps a distinctly bluish-green appearance, and results in extremely poor rendition of colors. Red objects, for example, usually look dull brown or even black under mercury lighting; human complexions take on an unfavorable hue. For this reason, mercury lamps are usually used in industrial lighting, street lighting, floodlighting or other applications where the appearance of colors is secondary to a desire for high efficiency from a relatively compact source.

Improvement in the color from mercury lamps is practical, and lamps of better color rendition are manufactured. These lamps usually employ a phosphor applied to the inner walls of the outer bulb. The phosphors used have the property of converting the ultraviolet energy from the arc to wavelengths not produced by the arc. Two basic types of phosphor-coated lamps are in common use. The "white" mercury lamps have a phosphor that produces blue-green, green, and yellow energy, but little or no red. These are higher in efficiency than ordinary mercury lamps, but still poor in over-all color rendition. The color-improved lamps use a phosphor that produces primarily red energy; the result is slightly lower efficiency but better rendition of human complexions and red colors. Even the color-improved mercury lamps, however, are regarded as being unsuitable for most commer-

are powerful, economical sources of ultraviolet energy. In addition to their common use as sun lamps, specially designed lamps are used in photochemical operations, such as the reproduction of tracings in blueprinting and black-and-white printing, and as black-light sources. The mercury arc radiates one wavelength in the near ultraviolet that is not harmful to humans; this energy is transmitted by the bulb glass, and has the property of causing some dyes and pigments—such as the whiteness added to many detergents—to fluoresce, or produce light. The mercury lamp can be used in a fixture covered with a filter that absorbs the light from the lamp but transmits the ultraviolet energy. The effects are frequently seen in striking decorations in dimly-lighted interiors, in some outdoor advertising, and in special-effects scenes in stage presentations and ice shows. See ULTRAVIOLET LAMP.

[A.MA.]

Mercury-vapor rectifier

age drop, high current capacity, and control characteristics, mercury-vapor tubes are widely used in rectifiers ranging in size from a few watts to thousands of kilowatts.

A *mercury-vapor tube* consists of an evacuated enclosure containing a number of electrodes. Conduction of electricity through the mercury vapor may take place between two or more of these electrodes. Mercury-vapor tubes may have either a heated cathode or a pool-type cathode. In both types, electric conduction through the mercury vapor takes the form of an arc discharge. In tubes of the hot-cathode type, electron emission is *thermionic* while in pool-cathode tubes, electron emission is produced by cathode spots. In both tubes, the voltage drop during conduction is low, because the electron space charge is neutralized by the ionized gas.

The terms mercury-vapor and mercury-arc are nearly synonymous and may be used interchange-

ably. In practice they sometimes have different meanings, mercury-vapor tube referring to tubes with a heated cathode containing only sufficient mercury to maintain the gas fill, and mercury-arc tube applying to tubes with a pool cathode of liquid mercury.

Hot-cathode mercury-vapor tube. Hot-cathode tubes have one or more anodes and control electrodes mounted in an evacuated enclosure containing mercury vapor. Those with only two electrodes are also known as phanotrons, while those with a control electrode between anode and cathode are called thyratrons. The most significant characteristic of the hot-cathode mercury-vapor tube is its low forward drop. Hot-cathode mercury-vapor tubes are widely used as rectifiers for the plate supply for radio transmitters and receivers. They also find many industrial uses in control and rectifier service for low- and medium-power applications.

Phanotron. In the smaller sizes, this mercury-vapor diode is similar in construction to the high-vacuum diodes used as rectifiers in radio receivers, except for the addition of a drop or two of liquid mercury. The operation of this tube is also similar to that of the high-vacuum diode, except that the anode current conduction does not begin until a starting voltage approximately equal to the tube drop (10–25 volts) is reached. Conduction also stops when the anode voltage drops below this characteristic tube drop.

The construction of a large size high-voltage tube with an indirectly heated cathode and a glass envelope is shown in Fig. 1. Some of the larger tubes designed for high current ratings in industrial applications feature a metal envelope with glass seals. Two-element mercury-vapor tubes range in size from 1–75 amperes at inverse voltages of 2000–22,000 volts. *See* GAS TUBE.

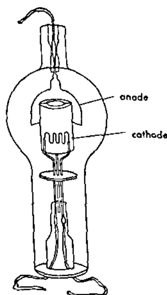


Fig. 1. Phanotron hot-cathode mercury-vapor tube.

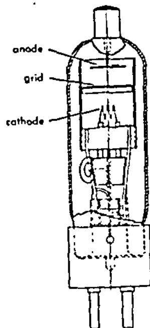


Fig 2. Thyratron with grid mounted between anode and cathode.

Thyratron. The construction of a typical three-element mercury-vapor glass thyratron with a filamentary cathode is shown in Fig. 2.

In the thyratron, the grid can control only the start of conduction. Before starting, when the tube is free of ions, it behaves like a high-vacuum tube, and the flow of current may be prevented by holding the grid at a potential below cut-off. When the grid voltage is raised slightly above cut-off with a positive potential on the anode, arc conduction develops rapidly and the grid has no further control. However, if the arc is interrupted by removal or reversal of anode voltage and time is allowed for the ions to diffuse to the walls, the process of starting may be repeated.

Thyratron tubes feature a high control ratio, because the grid requires little energy and can be controlled from a high impedance source. Ratings range from a few milliamperes to 18 amperes at inverse voltages of 500–25,000 volts. *See* THYRATRON.

Pool-cathode mercury-arc tube. This electron tube has a mercury-pool cathode, one or more anodes, and a number of auxiliary electrodes all enclosed in an evacuated envelope.

In addition to the low forward voltage drop which is characteristic of all mercury-vapor tubes, tubes with a pool cathode have a higher current capacity and longer life than the hot-cathode tubes because of the indestructible nature of the mercury-pool cathode. Pool-cathode mercury-arc tubes are widely used for medium and high-power applications in welding and rectifier service.

Arc conduction. In pool-cathode tubes conduction is initiated by establishing a cathode spot on the surface of the mercury pool. A cathode spot may be produced by applying a strong electric field

at the mercury surface and either passing current into the mercury from a movable electrode while that electrode is withdrawn or passing current into the mercury through a suitable ignitor extending into the mercury. The cathode spot is a source of the electrons, which ionize the mercury vapor and form the principal arc current.

Conduction between anode and cathode occurs when the anode voltage becomes positive. During conduction a current of electrons flows from cathode to anode through an oppositely directed stream of slowly moving positive ions. The positive ions neutralize the negative space charge produced by the electrons. As a result, the voltage drop across the arc is low, being just sufficient to supply the energy to ionize the mercury vapor and replace the ions lost by recombination at the walls.

At the end of conduction the anode current falls to zero and the mercury ions recombine with electrons to form neutral atoms. If a negative voltage is now applied to the anode with respect to the cathode, the tube will not conduct current because there is no source of electrons at the anode.

Pool-cathode tubes contain a large pool of liquid mercury. During operation, a blast of mercury vapor flows upward from the cathode spots on the mercury surface. This vapor is condensed on the tube walls and returns to the pool in liquid form. The mercury-vapor pressure in pool tubes is determined by the vapor pressure characteristics of mercury vapor and the temperature of the coolest part of the tube. Figure 3 shows the relation between vapor pressure and temperature for mercury. The vapor pressure doubles every 10 degrees in the range from 30–70°C. As a result, mercury-vapor tubes are greatly affected by temperature and must be operated within specified temperature limits. If temperature is too low, there is insufficient gas to neutralize the electron space charge, resulting in arc starvation and surges. At too high temperatures, voltage breakdown occurs during the inverse period.

Mercury-pool tubes are capable of operating at high currents without danger since the cathode pool is continuously replenished by condensation. They must, of course, be provided with adequate cooling surface to remove the arc losses and prevent a rise in temperature that would affect the gas pressure.

Control. The start of conduction in pool tubes may be controlled by either a grid or an ignitor, depending on the type of tube. In tubes of the multianode or excitron type, where a cathode spot is maintained continuously, a grid mounted near the anode is required to control the start of anode conduction. This grid functions in the same manner as it does in a hot-cathode tube. In tubes of the ignitron type, where the cathode spot is established each cycle and is permitted to go out at the end of anode conduction, control may be obtained by varying the time at which the ignitor is energized. *See* EXCITRON; IGNITRON.

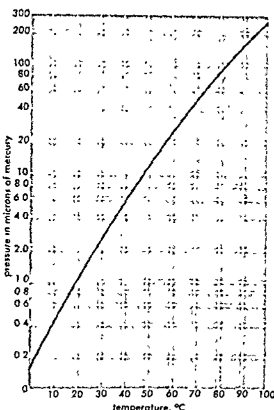


Fig. 3. Relation between mercury-vapor pressure and temperature for equilibrium conditions.

Phase control is the process of varying the point within the cycle at which anode conduction is permitted to begin. It provides a means of varying the output voltage of a rectifier. With free conduction (no phase control), the dc output voltage is a maximum, because each anode carries current during the part of the cycle when its phase voltage is highest. With phase control, the start of anode conduction is delayed and the dc output is reduced. The amount of phase control may be expressed in two ways: (1) the reduction in dc voltage obtained by phase control or (2) the angle of phase retard or advance. Grid or ignitor control may also be arranged to block conduction completely for switching or welding applications.

Arc-back. An arc-back (the flow of anode current in the reverse direction) is caused by the formation of a cathode spot on the anode. Arc-backs occur at infrequent intervals in mercury-arc tubes, even under normal operating conditions, and their occurrence is essentially random in nature. However, the frequency of arc-back depends upon the duty imposed on the tube and increases with increase in voltage, current, amount of phase control, and by such conditions of the rectifier tubes as temperature, internal cleanliness, and vacuum. If the capabilities of a tube are exceeded, structural failure usually does not occur immediately, but instead, the tube fails to rectify, that is, it fails by

arc-back. For this reason the maximum rating of a rectifier tube is usually fixed by its reliability as measured in terms of arc-back frequency. High-power rectifier tubes are generally designed to operate with a few arc-backs per year.

The effect of an arc-back in a rectifier unit is to apply a short circuit on both the ac and dc circuits. Since arc-backs are of random nature and it is not practical to eliminate them completely, rectifier equipment must be provided with protective switchgear to limit the effect of arc-back and prevent damage to the tubes and associated equipment.

Multianode tube. Pool tubes of the multianode type have been built in many forms. The smaller tubes are usually of sealed glass construction with 2, 3 or 6 anodes and air cooling. In the larger sizes, the tubes (or tanks) are made of metal and may have 6, 12, or 18 anodes. They may be of either pumped or sealed construction with either air or water cooling. Many types and sizes of multianode metal-tank mercury-arc rectifiers are installed and in operation in transportation, electrochemical, and general industrial service throughout the world. While multianode rectifiers have been superseded by single-anode types in the United States, they are still being manufactured in Europe in both glass and metal construction.

The construction of a typical pumped multianode metal-tank rectifier is shown by the sectional view in Fig 4.

In the multianode tank, cathode emission is established by means of the starting electrode, which consists of a metal rod mounted above the cathode pool so that it can be dipped into the mercury by means of a solenoid. When the starting electrode is withdrawn from the mercury pool, a low-voltage circuit is opened and a cathode spot is formed on

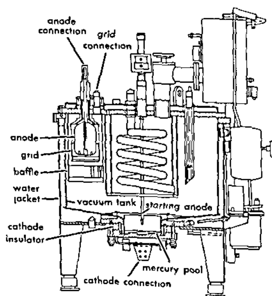


Fig. 4 Multianode, water-cooled, mercury-arc rectifier rated at 1500 kw, 600 volts. (General Electric Co.)

the mercury surface. The cathode spot is maintained continuously by the excitation anodes, which may be connected to either an ac or dc low-voltage source.

In any continuously excited metal tank, the metal envelope must be insulated from the cathode pool to prevent the transfer of cathode spots to the tank walls where they would melt the steel, contaminate the interior, and lead to permanent damage and failure.

Current ratings of glass tubes range up to 400 amperes and of metal tanks up to 8000 amperes per tank, while voltage ratings range from 250 to 20,000 volts. Rectifier equipment using multianode pool tubes usually ranges in size from 10 to 5000 kw units. [C.C.H.]

Bibliography: J. D. Cobine, *Gaseous Conductors*, 1958; W. G. Dow, *Fundamentals of Engineering Electronics*, 2d ed., 1952; A. E. Knowlton, *Standard Handbook for Electrical Engineers*, 9th ed., 1957; L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases*, 1947; O. K. Marti and H. Winograd, *Mercury Arc Power Rectifiers*, 1930; H. J. Reich, *Theory and Application of Electron Tubes*, 2d ed., 1944.

Meridian

That half of a great circle on Earth that passes through points having the same longitude and terminates at the North and South Poles. The meridian from which longitudes are measured is the one passing through Greenwich, near London, England; it is called the prime meridian.

The *celestial meridian* is a great circle on the celestial sphere passing through the two celestial poles and the observer's zenith. Two branches of it are distinguished, each extending from pole to pole, the upper branch containing the zenith and the lower the nadir. The meridian passage, or culmination, of a celestial object is its crossing over the celestial meridian, upper and lower culminations referring to the upper and lower branches. Stars nearer to the celestial pole than a distance equal to the latitude of the observer are above the horizon at both upper and lower culminations, and are called circumpolar stars. *See* ASTRONOMICAL COORDINATE SYSTEMS. [G.M.C.]

Meristem, apical

A region of embryonic tissue occurring at the tips of roots and stems. The numerous cell divisions that occur in apical meristems produce the cells that make up the primary plant body and are responsible for the growth in length, characteristic of most vascular plants. The self-perpetuating initial cells of apical meristems and their derivatives are not sharply delimited from each other. Therefore, the initials and some of their immediate derivatives are often jointly referred to as promeristem. Each of the terms shoot apex and root apex corresponds approximately to the promeristem. The partly differentiated but still meristematic regions

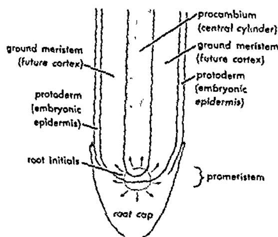


Fig. 1. Diagram of a root apical meristem. Cortex and central cylinder have separate initials, epidermis and root cap have a common origin.

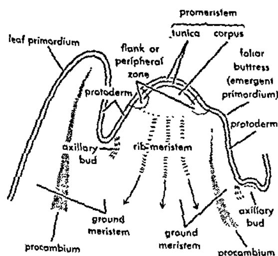


Fig. 2. Diagram of a hypothetical shoot apical meristem with a 2-layered tunica.

back of the promeristem are conveniently subdivided according to the tissue systems they produce (Figs. 1,2). Thus, the protoderm gives rise to the epidermal system; the procambium, to the primary vascular system; and the ground meristem, to the fundamental, or ground tissue system (pith, cortex). These three subdivisions are often called the primary meristems.

Root apical meristem. This area lies behind a protective structure, the root cap, and is, therefore, subterminal rather than strictly apical. It produces no lateral appendages. The structure of root tips is variable, and there is no constant relationship between the organization of the initial regions and the tissue systems they produce. In many vascular cryptogams (lower forms, such as *Equisetum* and leptosporangiate ferns) there is a prominent apical cell (Fig. 3*a*). In other plants there are several to many initials, and their delimitation under the microscope may be uncertain. In grasses the root cap is derived from independent initials (the calyptrogen) (Fig. 3*b*). Many gymnospermous roots have two sets of initials. One set produces the central (vascular) cylinder; the other produces the cortex and root cap. The outermost layer of the cortex becomes the epidermis. Another common type of organization occurs in certain dicotyledons (*Raphanus*, *Nicotiana*) where one tier of initials produces the central cylinder, a second tier forms the cortex, and a third tier the epidermis and root cap.

Branch roots. These arise at some distance from the apex and are formed from deep tissue regions (Fig. 4), commonly the pericycle. In some vascular cryptogams the endodermis is involved. Roots that arise on stems, leaves, and various organs other than roots are known as *adventitious roots*. They, too, ordinarily have an endogenous (internal) origin. The capacity to produce adventitious roots is of great practical importance in plant propagation, for example, in cuttings.

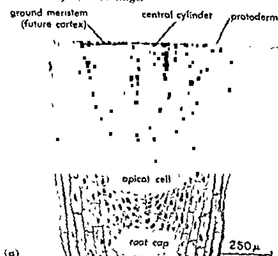


Fig. 3. Median longisections of apical meristems of roots. (a) Root tip of *Equisetum*; (b) Root tip of *Zea*.

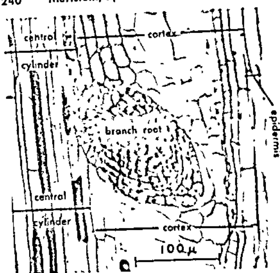


Fig 4. Longisection of part of root of *Eichhornia* showing endogenous origin of branch root primordium.

Shoot apical meristems. These are protected by young leaves, bud scales, and stipules, and the structure of the shoot apical meristems is complicated by the initiation of these organs. Both external form and size are variable. The apical cone (promeristem) is considerably elongated if the lateral appendages are formed well down its sides (*Elodea* and many grasses). Commonly it is more or less hemispherical. It may also be broad and nearly flat (certain cacti and many other angiosperms), or even depressed below the youngest primordia (*Drimys*). The diameter of shoot meristems varies from about 50 μ (*Zea*) to 3.5 mm (*Cycas revoluta*) In certain cacti, the meristems range up to 1.5 mm in diameter, the largest known for angiosperms. Apical meristems of shoots also vary in size and structure from seedling stages to maturity in the same plant. Moreover, most of them (except possibly the very largest) show changes in size and form during the period (plastochron) between the initiation of one lateral appendage and the next.

The internal architecture of shoot apices shows a range of variation comparable to that found in roots. Many lower vascular plants as well as nonvascular plants (mosses, liverworts, some red and brown algae) have apical cells (Fig. 5*a*). In the Marattiales and in *Lycopodium* there are usually several large initial cells (Fig. 5*b*). Gymnospermous and angiospermous shoot apices show a complex pattern of zonation resulting from differences in size and vacuolation of cells as well as differences in rate and direction of division. As in roots, there is no constant relationship between the zones and the tissue systems derived from them. Basically, there is a group of initials at the summit of the meristem. Lateral derivatives of the initials constitute a peripheral, or flank zone, in which leaf initiation occurs and in which cell division is more frequent than in the initials themselves. Close beneath the initials there may be a rib-meristem (files of cells which divide predominantly in planes perpendicular to the axis), which is the precursor of the pith (Fig. 5*c*). In most gymnosperms, periclinal (parallel with the circumference) divisions occur in the outermost layer of the shoot apex. *Ephedra* and *Gnetum* are exceptions to this rule. Angiosperms are characterized by having 1-5 layers over the summit of the shoot tip in which divisions are anticlinal (perpendicular to the surface) The layered zone is often called the tunica, while the central core covered by the tunica is called the corpus. Divisions occur in various planes in the corpus. The tunica-corpus pattern is superimposed upon other zones. For example, the flank zone of leaf initiation involves the tunica and variable amounts of the corpus, depending upon the species.

Primary thickening meristems. These are found in many monocotyledons which have relatively thick stems (palms, woody Liliales, and bulb-bearing plants) and in some dicotyledons (certain cacti) in which there is relatively little secondary growth. These meristems are usually not sharply delimited, but they tend to be most active in a zone beneath the leaf primordia and young leaves (Fig. 5*d*). Their derivatives often form files of cells which diverge outward and upward from the base of the apical meristem. The apical cone itself then lies in a depression which may be several centimeters deep.

Leaf primordia. These arise from localized areas of intensified cell division on the flanks of shoot apices. The emergent bulge represents the leaf base and is sometimes called a foliar buttress. The body of the leaf grows upward from the but-

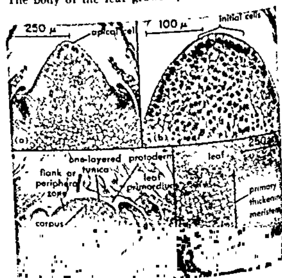


Fig. 5. Apical meristems of shoots. (a) Median ... section of shoot apical meristem of *Equisetum* showing apical cell. (b) Median longisection of apical meristem of *Lycopodium* showing several large initial cells. (c) Longisection of apical meristem of *Opuntia cylindrica* (d) Part of median longisection of shoot tip of *Yucca* showing primary thickening meristem. Apical meristem is about 2 mm to the left.

tress. The insertion of a leaf primordium may include a small sector of the apical meristem, or it may involve the entire circumference (grasses, *onion*) in which case the mature leaf has a sheathing base. The peglike structure which develops from the foliar buttress represents the axis (petiole and midrib) of the future leaf. In many ferns, the leaf undergoes a relatively long period of apical growth. In seed plants the period of apical growth is restricted to primordial stages and, after the apex matures, further growth of the leaf is intercalary. The leaf blade, or lamina, originates from localized regions of intensified meristematic activity on the margins of the primordium. If the blade meristem forms a continuous strip along each margin, it produces a simple leaf (Fig. 6*a*); if the marginal meristem is repeatedly interrupted, it forms a compound leaf. In peltate (shield-shaped) leaves the lower margins of the blade meristems are continuous across the adaxial (toward the axis) face of the leaf primordium.

Stem branch primordia. In contrast to root branch primordia, stem branch primordia usually have a superficial (exogenous) origin (Fig. 6*b*). In the majority of vascular plants they originate in or near the axils (angle between leaf and stem) of leaves and are known as axillary buds. Frequently axillary buds are initiated on the axis and subsequently displaced toward the leaf axil, or even onto the leaf itself. They may appear singly or in groups (serial or supernumerary buds). Although buds usually arise later than the subtending leaf primordium, the time of initiation is variable. Often they seem to arise directly from the apical meristem, in association with the youngest leaf primordia. They may also arise later, from partly differentiated tissues of the internode. In grasses the bud originates in close association with the leaf primordium above it and the two are later separated by the interpolation of an internode. Stem buds which arise elsewhere than in or near leaf axils, or on organs other than stems, are known as adventitious buds. The capacity to produce adventitious buds is utilized in plant propagation (*Begonia*, sweet potato, African violet). When a branch begins to grow, its apical meristem usually becomes similar to that of the parent shoot.

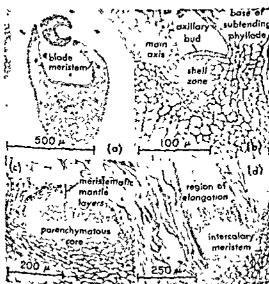


Fig. 6. Various meristems. (a) Transection of young leaf of *Pereskia aculeata* showing blade meristem. (b) Part of longisection of shoot tip of *Acacia longifolia* showing exogenous origin of axillary bud. (c) Longisection of early floral primordium of *Neobesseya*. (d) Longisection of base of spine of *Opuntia* showing intercalary meristem and region of elongation.

Among certain vascular cryptogams, branching occurs without relationship to the leaves. If the parent shoot forms two equal branches, the system is dichotomous (*Psilotum*, *Lycopodium*); if the branch arises at the side of the apical meristem, the system is monopodial. Dichotomous branching, though rare in seed plants, occurs in a few of them (*Mammillaria perbella*, *M. parkinsonii*).

Floral meristem. An apical meristem of a shoot that produces primordia of flower parts instead of leaves is called a floral meristem. Many plant anatomists regard floral meristems as differing from vegetative meristems only in degree. There are other schools of thought. In general, floral meristems are determinate and tend to grow in width rather than in length. Cells in the interior of a floral primordium usually become vacuolated early and show signs of becoming mature, whereas those on the periphery (the mantle layers) retain their meristematic characteristics longer and are involved in the initiation of the floral appendages (sepals, petals, stamens, and carpels) (Fig. 6*c*). In general, the initiation and early developmental stages of floral appendages are similar to those of leaves. Floral meristems may originate by the gradual transformation of a vegetative apical meristem, or they may arise in the axils of leaves or bracts, after the manner of an axillary bud.

Intercalary meristems. These are regions of residual meristematic activity which remain after differentiation has occurred between them and the apical meristem. Probably the best known examples are those that occur above the nodes and in the bases of leaf sheaths in grasses and above the nodes in young shoots of *Equisetum*. Striking examples of intercalary meristems in foliar organs are found at the bases of growing cactus spines (Fig. 6*d*) and at the bases of *Welwitschia* leaves. See CYTOLOGY; PLANT ORGANS; PLANT PHYSIOLOGY; PLANT TISSUE SYSTEMS. [N.H.B.]

Bibliography: See PLANT ANATOMY.

Meristem, lateral

Strips or cylinders of dividing cells located parallel to the long axis of the organ in which they occur. Radial enlargement of the cells derived from these meristems increases the diameter of the organ. The lateral meristem is concerned with secondary growth in the sense that its meristematic activity adds cells to the primary body which was

derived from the apical meristems (*see* MERISTEM, APICAL).

Vascular cambium. The vascular cambium produces secondary xylem on its inner face and secondary phloem on its outer face (*see* PHLOEM; XYLEM). In a woody plant the cambium occurs as a continuous sheath, or hollow cylinder, separating xylem and phloem. In stems with separate vascular bundles, fascicular cambium arises within the bundles as strips from the residual procambial cells lying between the primary xylem and primary phloem. In plants with uninterrupted cylinders of secondary vascular tissues, interfascicular cambia often arise in the ray parenchyma between the bundles in lateral connection with the fascicular cambial strips within the vascular bundles; thus the alternating strips of fascicular and interfascicular cambium form a complete cambial cylinder (Fig. 1). The vascular cambium in woody roots arises as strips on the inner faces of the primary phloem groups and it becomes a laterally continuous sheath when the pericycle cells outside the primary xylem ridges become transformed into cambial initials. *See* ROOT (BOTANY).

The vascular cambium has two kinds of initial cells, ray and fusiform, or spindle-shaped (Fig. 2). Because the immediate derivatives of these cambial cells are quite similar to the initials and may undergo further divisions, these derivatives and the initials together are said to constitute the cambial zone. The zone may be several cell layers in radial dimension. The composition of the cambium itself, [illegible] creasing girth of an organ by self-multiplication of the fusiform initials and by the origin of entirely new groups of ray initials.

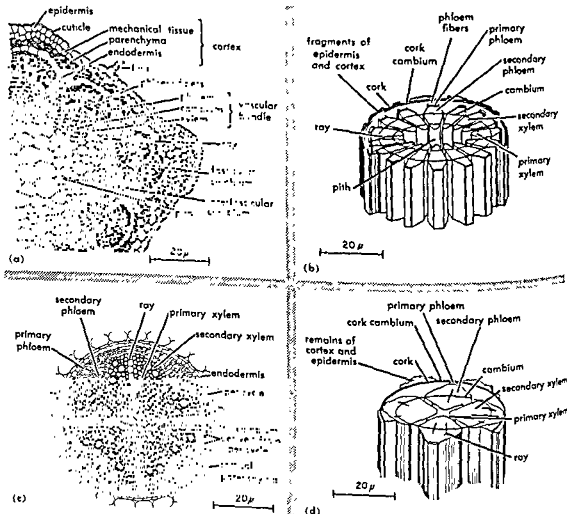


Fig. 1. Diagrams of stem and root showing lateral meristems, vascular cambium, and cork cambium. (a) Part of a cross section of an alfalfa stem. (b) Three-dimensional diagram of a dicotyledonous stem, showing the results of cambial activity. (c) Cross section of the vascular cylinder of a dicotyledonous root at a level at which the cambium has become a continuous layer and has produced considerable amounts of secondary tissues, diagrammatic. (d) Three-dimensional diagram of a part of a root where there has been considerable secondary thickening. (*From G. M. Smith et al., Textbook of General Botany, 5th ed., Macmillan, 1953*)

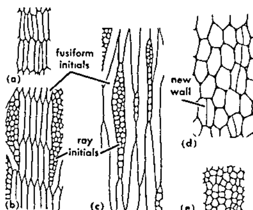


Fig. 2. Tangential views of cambium. (a) Root, *Bougainvillea spectabilis*, cork cambium with but one type of initial, elongate polygonal in shape. (b) Stem, *Robinia pseudoacacia*, storied vascular cambium with ray and short fusiform initials. (c) Stem, *Juglans regia*, nonstoried vascular cambium with ray and long fusiform initials. (d) Stem, *Yucca gloriosa* (a monocotyledon), unusual vascular cambium with but one type of initial. (e) Stem, *Lindera benzoin*, cork cambium with but one type of initial, nearly isodiametric in shape

Fusiform initials. These are vertically elongate cells, brick-shaped in transectional view, with tapering ends, and with an extensive vacuolar system. These initials give rise to all cells in the vertical system of secondary xylem and phloem. As seen in tangential views, storied cambium has an orderly arrangement in horizontal tiers of uniformly short fusiform initials whose ends slightly overlap those of initials in juxtaposed tiers. The more primitive and less orderly nonstoried cambium lacks obvious tiers, and has longer initials of variable length whose ends greatly overlap those of other initials. The radial walls of fusiform initials are relatively thick (with many depressed primary pit fields) in the dormant season, but become much thinner during part of the growing season when the cells are actively dividing. Divisions in the tangential (periclinal) plane produce secondary tissues. Multiplication of initials—which may occur in the dormant season—results from radial (anticlinal) divisions. The separating walls in these divisions may be highly oblique to almost transverse in long initials, or approximately vertical in short initials. The derivatives of oblique divisions elongate by apical intrusive growth.

Ray initials. These cells are nearly isodiametric in form and, though much shorter vertically, often have a wider radial dimension than do fusiform initials. Ray initials divide periclinally in producing new phloem and xylem ray cells. They may also divide anticlinally or transversely and thus add to the number of initials in a group. An entirely new group of ray initials may have its origin in a segment cut off from the end of a fusiform initial or through subdivision of an entire fusiform initial.

Cork cambium. This cambium (phellogen) produces cork (phellem) on its outer face and sometimes phelloderm (parenchyma) on its inner face (*see* PERIDERM). The initials, unlike those in the vascular cambium, are of a single type. They are short, appear rectangular and flattened in transectional and radial views, and rectangular to polygonal in tangential aspect. Cork and phelloderm are formed by periclinal divisions. The increase in the circumference of the cambium occurs by anticlinal divisions. Phellogen initials in certain monocotyledons divide only a few times and are then transformed into cork cells themselves.

Other cambia. The initials responsible for production of secondary tissues in certain woody mon-

ondary vascular bundles, but strictly periclinally in forming secondary parenchyma.

Some vascular cambia differ from the ordinary type chiefly in their unorthodox activity. For example, several concentrically oriented cambia are simultaneously active in the garden beet. Cambia in some species may remain active for a time and then become transformed into vascular tissue. New cambium then originates outside the secondary phloem. Some cambia add both xylem and phloem to the inner side. *See* PLANT TISSUE SYSTEMS.

[V.I.C.]

Bibliography: *See* PLANT ANATOMY.

Mermithoidea

Usually considered a superfamily of nematodes, this group contains two families: the Tetradonematidae, which are parasites all their life, and the more common Mermithidae, which are parasites only as larvae. Some authorities consider the Mermithoidea to be an order. The Mermithidae, filiform nematodes often 20 cm long, have the

snails, are infected by eating eggs or by the attack of infective larvae. The larvae grow in the host's hemocoele and emerge from insects usually prior to host pupation, at which time the host usually dies. Maturation, mating, and oviposition occur in soil and fresh water. Females ovipositing on grass are often mistaken for hairworms, but they are white whereas hairworms are brown. *See* NEMATODA; NEMATOMORPHA.

[H.E.W.]

Merogony

The normal or abnormal development of a part of an egg, following cutting, shaking, or centrifugation of the egg before or after fertilization. The primary importance of merogony lies in the detection of often invisible, regional differences in the cytoplasm of an egg which later manifest themselves by the appearance of typical abnormalities in the embryo or larva.

Depending on the character of the nucleus present in the egg fragment, the following types of

merogony are recognized: (1) diploid merogony, the nucleus is the normal fusion product of the egg and sperm nuclei and thus diploid; (2) andromerogony, development with the haploid sperm nucleus; (3) gynomerogony, development of a fragment of a fertilized egg with the haploid egg nucleus; (4) parthenomerogony, development of a nucleated fragment of an unfertilized egg following parthenogenetic stimulation. *See* ANDROGENESIS; EMBRYOLOGY, EXPERIMENTAL; FERTILIZATION; GYNOGENESIS; PARTHENOGENESIS. [G.FA.]

Merostomata

A class of primitive arthropods of the subphylum Chelicerata. These chelicerates are distinguished by their aquatic mode of life and the possession of certain abdominal appendages which bear respiratory organs. Merostomes are known from as early as Lower Cambrian. Three species are known today, including the king crab, *Limulus*. Merostomes have in common (1) a chitinous to phosphatic exoskeleton, comprising a cephalothorax and an abdomen with telson; (2) six uniramous, stenopodous cephalothoracic appendages, of which the first pair is chelicerate, and the remainder are walking legs, with masticatory gnathobases; (3) compound eyes, except in Synxiphosura. *See* CHELICERATA.

Classification of the Merostomata varies. Some authorities consider the member groups as orders, others treat them as subclasses, orders, and suborders, as shown in the two following classifications:

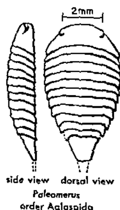
Phylum Arthropoda
  Class Arachnoidea
    Subclass Merostomata
      Order Aglaspida
      Order Eurypterida
      Order Xiphosuria
      Order Synxiphosura
      Order Limulina
Phylum Arthropoda
  Subphylum Chelicerata
    Class Merostomata
      Subclass Xiphosaura
        Order Aglaspida
        Order Xiphosurida
          Suborder Synxiphosaurina
          Suborder Limulina
      Subclass Eurypterida

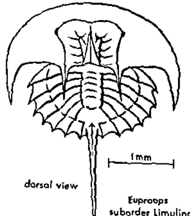
eyes, unspecialized cephalothoracic appendages (the chelicerate first pair excepted), a preabdomen with unspecialized walking legs, twelve freely articulating abdominal segments (somites), and a styliform telson united to the last segment.

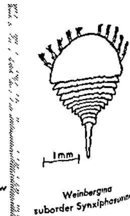
The Eurypterida (Ordovician to Permian) show more specialization of the cephalothoracic appendages, gill books replace walking legs on the preabdomen, well-developed genitalia occur ventrally on the second abdominal somite, the postabdomen has ventral sternites joined with dorsal tergites to produce telescoping segmental rings, and the telson (spiniform to scutiform) is freely articulated with the terminal segment. *See* EURYPTERIDA.

The Xiphosura (Early Devonian to the present) retain most of the specializations of the Eurypterida, such as cephalothoracic appendages, ocelli (visual organs), genitalia, and gill books. They also retain some of the more primitive, aglaspid characters, such as a somewhat trilobed body form and freely articulating styliform telson. Moreover, the original twelve abdominal segments are reduced to from six to nine, and are partially or wholly ankylosed (fused). The living fossil, *Limulus*, is included in this group. *See* XIPHOSURA; *see also* LIVING FOSSILS.

The Synxiphosura (late Silurian to early Devonian) have much in common with the primitive Aglaspida. This heterogenous group may represent an explosive proliferation of aberrant, terminal and apparently blind forms.







Typical genera of class *Merostomata*. (*After* R. C. Moore (ed.), *Treatise on Invertebrate Paleontology*, Geological Society of America, 1955)

The Aglaspida in the early Paleozoic and the Xiphosura in the post-Paleozoic inhabited normal marine, shallow benthonic environments, but the Eurypterida, Synxiphosura, and primitive Xiphosura of Middle and Late Paleozoic preferred either fresh or abnormally saline habitats.

The closest relationship of the Merostomata seems to be with the scorpions via the Paleoscorpionida. [C.O.R.]

Bibliography: R. C. Moore (ed.), *Treatise on Invertebrate Paleontology*, 1955; R. C. Moore, C. G. Lalicker, and A. C. Fischer, *Invertebrate Fossils*, 1952; R. R. Shrock and W. H. Twenhofel, *Principles of Invertebrate Paleontology*, 2d ed., 1953.

Merriam's life zones

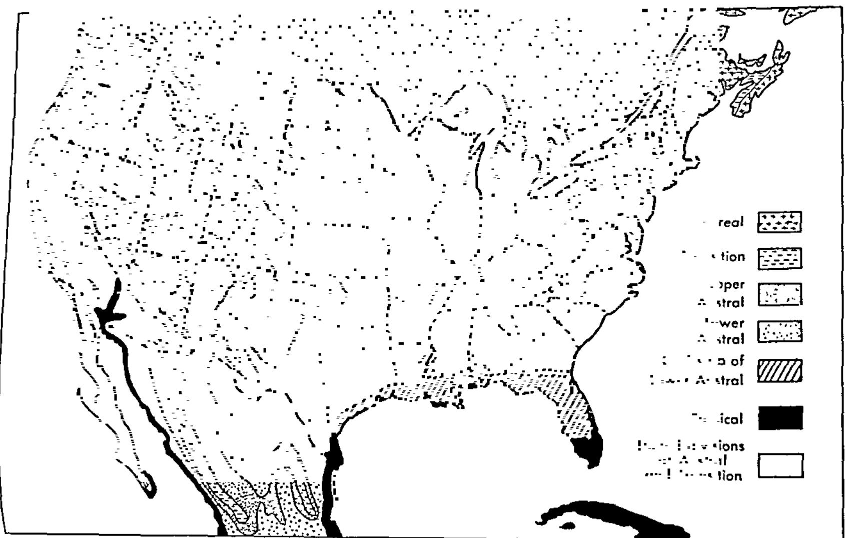
Broad climatic belts across the North American continent designed to contain the habitats of America's important animal forms. Although the change in climate is gradual, the life zones are defined so that each one has a character of its own. Life zones correlate reasonably well with major crop regions and to some extent with general vegetation types.

The life zones were the results of surveys and explorations by C. Hart Merriam and other zoologists of the U.S. Biological Survey. At the time, the late nineteenth century, vegetation maps and classifications were not available. Later studies of the vegetation by plant ecologists have emphasized different criteria and made it necessary to revise Merriam's boundaries. However, Merriam's work provided the initial stimulus and much of the terminology he proposed persists, especially in zoogeographic literature in North America.

Isotherms based on sums of effective temperatures correlated with observed distributions of certain animals and plants led to Merriam's first law that animals and plants are restricted in northward distribution by the total quantity of heat during the season of growth and reproduction. The

Characteristics of Merriam's life zones

Zone name	Example	Vegetation	Typical and important plants	Typical and important animals	Typical and important crops
Arctic-alpine zone	Northern Alaska, Baffin Island	Tundra	Dwarf willow, lichens, heathers	Arctic fox, muskox, ptarmigan	None
Hudsonian zone	Labrador, southern Alaska	Taiga, coniferous forest	Spruce, lichens	Moose, woodland caribou, mountain goat	None
Canadian zone	Northern Maine, northern Michigan	Coniferous forest	Spruce, fir, aspen, red and jack pine	Lynx, porcupine, Canada jay	Blueberries
Western division					
Humid transition zone	Northern California coast	Mixed coniferous forest	Redwood, sugar pine, maples	Blacktail deer, Townsend chipmunk, Oregon ruffed grouse	Wheat, oats, apples, pears, Irish potatoes
Arid transition zone	North Dakota	Conifer, woodland sagebrush	Douglas-fir, lodgepole, yellow pine, sage	Mule deer, whitetail, jackrabbit, Columbia ground squirrel	Wheat, oats, corn
Upper Sonoran zone	Nebraska, southern Idaho	Piñon, savanna, prairie	Junipers, piñons, grama grass, bluestem	Prairie dog, blacktail jackrabbit, sage hen	Wheat, corn, alfalfa, sweet potatoes
Lower Sonoran zone	Southern Arizona	Desert	Cactus, agave, creosote bush, mesquite	Desert fox, four-toed kangaroo rats, roadrunner	Dates, figs, almonds
Eastern division					
Alleghenian zone	New England	Mixed conifer and hardwoods	Hemlock, white pine, paper birch	New England cottontail rabbit, wood thrush, bobwhite	Wheat, oats, corn, apples, Irish potatoes
Carolinian zone	Delaware, Indiana	Deciduous forest	Oaks, hickory, tulip tree, redbud	Opossum, fox, squirrel, cardinal	Corn, grapes, cherries, tobacco, sweet potatoes
Austroriparian zone	Carolina piedmont, Mississippi	Long-needle conifer forest	Loblolly, slash pine, live oak	Rice rat, woodrat, mocking bird	Tobacco, cotton, peaches, corn
Tropical zone	Southern Florida	Broadleaf evergreen forest	Palms, mangrove	Armadillo, alligator, roseate spoonbill	Citrus fruit, avocado, banana



Life zones of the United States. (C. H. Merriam, 1898)

mean temperature for the six hottest weeks of the summer formed the basis for the second law that animals and plants are restricted in southward distribution by the mean temperature of a brief period covering the hottest part of the year.

In Canada the life zones are entirely transcontinental. Because of faunistic and climatic differences, the eastern and western parts of most life zones in the United States had to be characterized separately. The western zones had to be further subdivided into humid coastal subzones and arid inland ones.

One of the concepts emphasized in the life-zone system is the correspondence in life forms between arctic and alpine areas, and between boreal and montane regions. It was recognized that latitudinal climatic zones have parallels in the altitudinal belts on mountain slopes, and that there is a faunistic affinity between such areas of similar temperature regime. [K.LE.]

Bibliography: E. L. Braun, *Deciduous Forests of Eastern North America*, 1950; J. R. Carpenter, *An Ecological Glossary*, 1956; C. H. Merriam, *Life Zones and Crop Zones of the United States*, USDA Bull. 10, 1898; J. E. Weaver and F. E. Clements, *Plant Ecology*, 2d ed., 1938.

Merwinite

A rare nesosilicate mineral with composition $Ca_3MgSi_2O_8$, crystallizing in the monoclinic system. Well-formed crystals of merwinite have not been observed; the mineral occurs in granular aggregates showing polysynthetic twinning. It is colorless or pale green with a vitreous luster. The hardness is 6 on Mohs scale, and the specific gravity is 3.15. Merwinite was first described in 1921 from Crestmore, near Riverside, California, where it was found intimately associated with gehlenite, spurrite, and monticellite. It has also been noted with spurrite and larnite at Scawt Hill, County Antrim, Ireland, and at Durango, Mexico. See SILICATE MINERALS. [C.S.HU.]

Meson

Any elementary particle with mass between those of the electron and the proton. The possibility of such particles was first raised in 1935 by H. Yukawa, who pointed out that the main features of nuclear forces would be explained if these forces were transmitted between nucleons through an intermediate field strongly coupled with nucleons, whose quanta were massive [200–300 electron masses (m_e)] and could carry electric charge between the nucleons.

In 1937, C. Anderson established the existence of positive and negative μ-mesons (muons) of mass 207m_e in cosmic radiation, but these were found to have a very weak coupling with nucleons. The mesons envisaged by Yukawa turned out to be the π-mesons (pions) of mean mass 270m_e, first identified in cosmic radiation by C. F. Powell (1947). Positive, negative, and neutral pions are produced copiously in nuclear collisions of sufficiently high energy, and their properties have been studied intensively, because of the availability of pion beams from various types of high-energy particle accelerators.

Soon after discovery of the pion, the existence of various K-mesons, with masses about 967m_e, was established by a number of cosmic ray researchers. Their detailed study has developed slowly, partly because their artificial production required the development of multi-Bev proton accelerators and partly because they are produced somewhat less copiously (at most, several per cent) than are pions at all proton energies available to date.

Decay processes and lifetimes. Positive and negative muons have essentially the same mean lifetime (2.2×10^{-6} sec), their dominant decay mode being to an electron or positron, a neutrino (ν), and an antineutrino ($\bar{\nu}$):

$$\mu^+ \rightarrow e^+ + \nu + \bar{\nu} \quad (1)$$

Positrons emitted from a spinning (polarized) μ^+-meson are found to have a spatial distribution asymmetric forward and backward relative to the μ^+-spin axis. Also, positrons (electrons) emitted in unpolarized $\mu^+(\mu^-)$ decay always spin in a sense right-handed (left-handed) with respect to their direction of motion. Either of these facts is sufficient to establish the failure of the parity conservation principle in the muon decay process. See PARITY (QUANTUM MECHANICS); SPIN (QUANTUM MECHANICS).

Positive and negative pions have the same mean lifetime (2.6×10^{-8} sec), their dominant decay mode being

$$\pi^+ \rightarrow \mu^+ + \nu \qquad \pi^- \rightarrow \mu^- + \bar{\nu} \quad (2)$$

(see Fig. 1). The muon emitted in $\pi^+(\pi^-)$ decay spins with a right-handed (left-handed) sense relative to its direction of emission, showing that parity conservation fails for the π-μ–decay processes. An alternate decay mode $\pi^+ \rightarrow e^+ + \nu$, with relative frequency 1.2×10^{-4}, has also been established for the positive pion.

The neutral pion decays rather rapidly to two photons, with a lifetime between 10^{-16} and 10^{-19} sec, $\pi^0 \rightarrow \gamma + \gamma$. It is believed that parity conservation holds in this process. The visible decay mode $\pi^0 \rightarrow \gamma + e^+ + e^-$ is well known and results from the internal conversion of one photon to an electron-positron pair; its relative frequency is 1.2×10^{-2}. The double internal conversion process $\pi^0 \rightarrow e^+ + e^- + e^+ + e^-$ has also become known, with relative frequency 3.5×10^{-5}.

The charged K-mesons have a considerable number of competing decay modes. The relative frequencies for corresponding decay modes are the same for K^+- and K^--mesons, which also have the same mean lifetime (1.2×10^{-8} sec). The main decay modes established for K^+- and K^--mesons, and (approximately) their relative frequencies, are as follows:

$K_{\mu2}$-mode (57%)
$$K^+ \rightarrow \mu^+ + \nu \qquad K^- \rightarrow \mu^- + \bar{\nu} \quad (3)$$

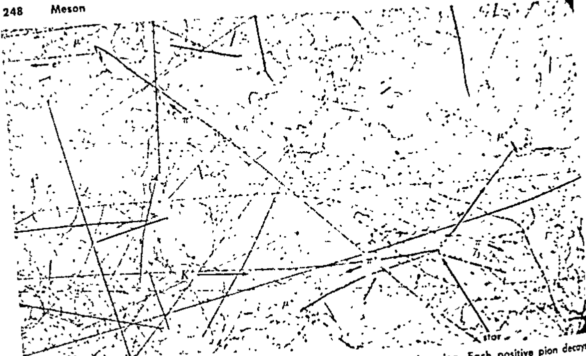


Fig. 1 Meson tracks made visible as strings of tiny bubbles in a chamber of liquid propane; photograph made while chamber was exposed to a beam of K^+ mesons from the Brookhaven Cosmotron. As indicated in the drawing, one series of tracks shows a K^+-meson entering from left and decaying into two positive pions and a negative pion. Each positive pion decays into a positive muon, which decays into an electron. Another pion enters the chamber from lower right and decays into a μ-meson at upper left. Experiment performed by D. A. Glaser and colleagues of the University of Michigan. (D. A. Glaser)

$K_{\pi 2}$(or θ)-mode (25%) $\quad K^+ \rightarrow \pi^+ + \pi^0$
$$K^- \rightarrow \pi^- + \pi^0 \qquad (4)$$

τ-mode (6%) $\quad K^+ \rightarrow \pi^+ + \pi^- + \pi^+$
$$K^- \rightarrow \pi^- + \pi^+ + \pi^- \qquad (5)$$

τ'-mode (2%) $\quad K^+ \rightarrow \pi^+ + \pi^0 + \pi^0$
$$K^- \rightarrow \pi^- + \pi^0 + \pi^0 \qquad (6)$$

$K_{\mu 3}$-mode (5%) $\quad K^+ \rightarrow \mu^+ + \nu + \pi^0$
$$K^- \rightarrow \mu^- + \bar{\nu} + \pi^0 \qquad (7)$$

K_{e3}-mode (5%) $\quad K^+ \rightarrow e^+ + \nu + \pi^0$
$$K^- \rightarrow e^- + \bar{\nu} + \pi^0 \qquad (8)$$

[illegible] and θ-modes led to [illegible]

that there existed no evidence that parity conservation was ever the case for any meson decay processes, and suggested experiments which then demonstrated the failure of parity conservation for decay processes in a very clear way. For K^+-decay, this has been illustrated explicitly by the observation that the muon emitted in the $K_{\mu 2}{}^+$-mode spins in a right-handed sense relative to its direction of emission.

The $K_1{}^0$ (or $\theta_1{}^0$)-meson is short-lived in comparison with the charged K-mesons, its lifetime being 1.0×10^{-10} sec. It has two main decay modes

$$K_1{}^0 \rightarrow \begin{cases} \pi^+ + \pi^- & (70\%) \\ \pi^0 + \pi^0 & (30\%) \end{cases} \qquad (9)$$

The $K_2{}^0$-meson has a much longer lifetime, about 6×10^{-8} sec. A number of decay modes,

$$K_2{}^0 \rightarrow \begin{cases} \pi^- + \mu^+ + \nu \\ \pi^+ + \mu^- + \bar{\nu} \\ \pi^- + e^+ + \nu \\ \pi^+ + e^- + \bar{\nu} \\ \pi^+ + \pi^- + \pi^0 \end{cases} \qquad (10)$$

have been identified for $K_2{}^0$-decay, but their relative frequencies are not yet well known.

Meson production processes. Pions are produced copiously in the collisions of nucleons of sufficiently high energy with nuclei, for example, through reactions such as

$$p + p \rightarrow \begin{cases} \pi^+ + d \\ \pi^+ + n + p \\ \pi^0 + p + p \end{cases} \qquad (11)$$

As the incident nucleon energy increases, the average multiplicity of pion production per collision increases, the production of two pions being most probable for energies of order 2000 Mev. At the highest cosmic-ray energies ($\gtrsim 10^{13}$ ev), the production of over 100 pions has been observed in a single collision. Pions are also produced in the collisions of pions, of photons, or of electrons of sufficiently high energy with nucleons. For example, the pion photoproduction processes

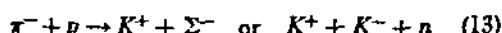
$$\gamma + p \rightarrow \begin{cases} \pi^0 + p \\ \pi^+ + n \end{cases} \qquad (12)$$

Table 1. Strangeness quantum numbers (s) for mesons and baryons

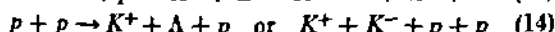
Particle	μ^+, μ^-	π^-, π^0, π^+	K^+	K_1^0	K_2^0	K^-	n, p	Λ	$\Sigma^-, \Sigma^0, \Sigma^+$	Ξ^-, Ξ^0
s	Not relevant	0	+1	+1 and −1	+1 and −1	−1	0	−1	−1	−2

are well known. Strong broad resonances (with half-widths of order 100 Mev) have been found in this process at photon energies of about 300 Mev, of about 750 Mev, and of about 1050 Mev.

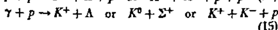
For K-mesons, the production processes are found to follow a rule of associated production, which may be simply stated if each particle with strong interactions is assigned the strangeness quantum number s given in Table 1. All strong reactions observed are then characterized by the law of conservation of strangeness, which states that the sum of the strangeness numbers s for all particles is the same after the reaction as before. For example, the following production reactions satisfy this conservation law:



$$\pi^- + p \rightarrow K^+ + \Sigma^- \quad \text{or} \quad K^+ + K^- + n \qquad (13)$$



$$p + p \rightarrow K^+ + \Lambda + p \quad \text{or} \quad K^+ + K^- + p + p \qquad (14)$$

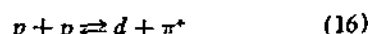


$$\gamma + p \rightarrow K^+ + \Lambda \quad \text{or} \quad K^0 + \Sigma^+ \quad \text{or} \quad K^+ + K^- + p \qquad (15)$$

It must be emphasized that this conservation law is not obeyed by the weak interactions which give rise to the K-meson decay processes (4), (5), (6), and (9) and to the hyperon decay processes (for example, $\Lambda \rightarrow \pi^- + p$). *See* HYPERON.

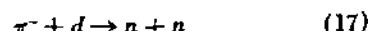
The muons observed in cosmic radiation or in accelerator beams mainly originate from the π-μ-decay process (2) following pion production processes. The direct electromagnetic process of μ^+-μ^- pair production by photons, the analog of the electron-positron pair creation process, has been established, but contributes very little to muon production in most circumstances.

Intrinsic spin and parity. The pion is known to be a pseudoscalar particle, that is, to be spinless with odd parity. The spin was first deduced from a comparison of the two reactions



$$p + p \rightleftarrows d + \pi^+ \qquad (16)$$

The parity was then deduced from the existence of the capture reaction



$$\pi^- + d \rightarrow n + n \qquad (17)$$

for π^--mesons which come to rest in deuterium.

At least three pieces of evidence indicate that the K-mesons are also spinless. These are the characteristics of the τ-mode of K^+-decay, the fact that the muons emitted in the $K_{\mu 2}{}^+$-decay mode and in π^+-decay [process (2)] have identical polarization, and the isotropic angular distribution observed for $K_1{}^0$-decay following all processes of $K_1{}^0$ production examined. Only the relative parities between K-mesons and Λ- and Σ-hyperons are physically meaningful. It is therefore conventional to assign even (+1) parity to the Λ-hyperon and to specify the parities of the K-meson and Σ-hyperons relative to this. However, this K-meson parity has not yet been established.

The muon has spin $\frac{1}{2}\hbar$ ($\hbar$ is Planck's constant h divided by 2π) and a magnetic moment of 8.8893 ± 0.0031 nuclear magnetons. This muon spin assignment follows from its magnetic moment, just the value expected for a spin $\frac{1}{2}\hbar$ particle without strong interactions; the electron energy distributions observed from the decay of polarized muons; and the agreement found between its electromagnetic processes (for example, muon-pair creation, bremsstrahlung by muons, and so on) and those predicted for a charged particle of spin $\frac{1}{2}\hbar$. The μ^+- and μ^--mesons have opposite parity, but the assignment of absolute parities to the muons would be without physical consequences.

Nuclear interactions. As illustrated in Fig. 2, all known particles with strong nuclear interactions, the pions and K-mesons as well as the nucleons and hyperons, form groups of "charge multiplets." The particles of each multiplet have almost the same mass but have different electric charges, the charge values increasing in unit steps of the positron charge across the multiplet (reading from left to right on Fig. 2). A particle multiplet including $2T + 1$ charge states is conveniently characterized

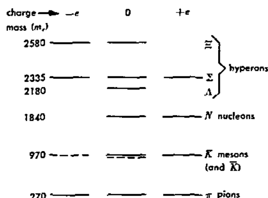


Fig. 2. The known "charge multiplets" of particles with strong nuclear interactions. Each particle is denoted by a horizontal line in the column appropriate to its electric charge, at a level corresponding to its mass value. The mass values of these multiplets are given in round numbers at the left. For the K-meson group, the doublet of anti-K particles ($\bar{K}$) is shown by dotted lines. There are also anti-particle multiplets corresponding to the nucleon doublet and to each hyperon multiplet; these are not shown on this diagram.

Table 2. The intrinsic properties of mesons*

Meson	μ^+	μ^-	π^+, π^-	π^0	K^+, K^-	K_1^0	K_2^0
Mass, m_e	206.9 ± 0.1		273.3 ± 0.1	264.3 ± 0.3	966.8 ± 0.4	974.4 ± 1.2	
Mean life-time τ, sec	$2.22 \pm 0.01 \times 10^{-6}$		$2.56 \pm 0.05 \times 10^{-8}$	$10^{-15} > \tau > 10^{-17}$	$1.224 \pm 0.013 \times 10^{-8}$	$1.00 \pm 0.04 \times 10^{-10}$	$(6.1^{+1.6}_{-1.1}) \times 10^{-8}$
Spin	½		0		0		
Parity†	$-p$	$+p$	odd (−1)		?		
Isotopic spin	Not relevant		1		½; the K-mesons consist of two doublets, (K^0, K^+) and their antiparticles (K^-, $\bar{K}^0$)		

* See also Table 1 for the strangeness quantum numbers s for mesons.
† The K-meson parity is defined with the convention of even parity (+1) for the Λ-hyperon. The quantity p for the μ-mesons is in principle not measurable, although the relative parity between μ^+ and μ^- mesons is known to be odd.

by the number T, called the isotopic spin of the multiplet (*see* ISOTOPIC SPIN). The values of T to be assigned to the meson multiplets are given in Table 2; thus, $T = 1$ for the pions, whereas the K-mesons consist of two $T = ½$ doublets, the (K^+,K^0) and ($\bar{K}^0$,K^-) doublets, where K^0 and $\bar{K}^0$ denote certain combinations of the K_1^0 and K_2^0 states which cannot be described in detail here.

This situation suggested to M. Gell-Mann the hypothesis that all strong nuclear interactions are charge-independent, that they depend on which multiplets the interacting particles belong to but not on the particular charges carried by these particles. For the nucleon-nucleon and pion-nucleon interactions, this principle of charge independence has been well tested and established, and it has been tested successfully for several K-meson reactions.

Pion scattering by protons has been measured for pion energies up to 4500 Mev. Its outstanding feature is the existence of a resonance strongly excited by pions of 195 Mev, the same resonance as that observed in the pion photoproduction process for 300-Mev photons. This resonant state is excited by incident pions carrying one unit ($\hbar$) of orbital angular momentum, and has total angular momentum $\frac{3}{2}\hbar$ and total isotopic spin $T = \frac{3}{2}$; it is generally referred to as the (3,3) isobar state. The two higher resonances observed in the pion photoproduction processes (12) are also excited by negative pions of ~600 Mev and ~900 Mev, respectively. These resonant states have isotopic spin $T = ½$; the lower state is excited by incident pions carrying two units of orbital angular momentum and has total angular momentum $\frac{3}{2}\hbar$, while the upper state has angular momentum $\frac{5}{2}\hbar$ and appears to be excited by incident pions carrying three units of orbital angular momentum. A further resonant state, with isotopic spin $T = \frac{3}{2}$, stands out clearly for positive pion scattering at about 1400 Mev; its further parameters are not yet known.

The strength of the pion-nucleon interaction may be measured by the coupling constant g, whose dimensions are those of electric charge and which plays a role in pion physics analogous to that played by the constant e (the electron charge) in electrodynamics. Consistent estimates for the dimensionless parameter $g^2/\hbar c$ (where c denotes the velocity of light) are obtained from the analysis of pion-nucleon scattering data, of pion photoproduction data, and of nuclear forces. These estimates of $g^2/\hbar c$ all lie close to 13.5; comparison of this value with $e^2/\hbar c \simeq 1/137$, the corresponding dimensionless constant of electrodynamics, provides an indication of the great strength of the pion-nucleon interaction.

The nuclear interaction of K^+- and K^--mesons has been examined up to about 2000 Mev kinetic energy. Up to 300 Mev, the K^+-nucleon processes are predominantly those of elastic scattering (including the charge-exchange process $K^+ + n \rightarrow K^0 + p$); in this region the K^+-proton nuclear interaction appears repulsive and energy-independent, whereas the K^+-neutron interaction increases with the K^+ energy. At higher energies, the K^+-nucleon interaction probabilities have relatively little dependence on K^+ energy. On the other hand, the K^--proton interactions consist of strong absorptive processes [see reaction (20a)] from which a pion and a Λ- or Σ-hyperon emerge, as well as strong elastic scattering. For sufficiently low K-meson velocities, the interaction probability for these absorptive processes becomes inversely proportional to velocity and therefore very large.

This contrast between the K^+ and K^- nuclear interactions is one reason for the assignment of opposite strangeness quantum numbers s to these two mesons. This is in accord with Gell-Mann's interpretation of the K^--meson as the antiparticle to the K^+-meson, as is also the close relationship between K^+- and K^--decay modes discussed earlier. More generally, the known K-mesons are now interpreted as a (K^+,K^0) doublet and its antiparticle doublet (K^-,$\bar{K}^0$).

For the muons, there exists no convincing evidence for any moderately strong nuclear interaction. Their observed scattering appears to be due only to the electromagnetic interactions arising from their charge and magnetic moment. The μ^--meson is the antiparticle to the μ^+-meson, as the μ^+-μ^--pair creation process illustrates. As a result, it is natural to regard the muon as a "heavy electron."

Nuclear capture; meson stars. A negative meson which comes to rest in matter is generally captured by an atom and then moves in an orbit

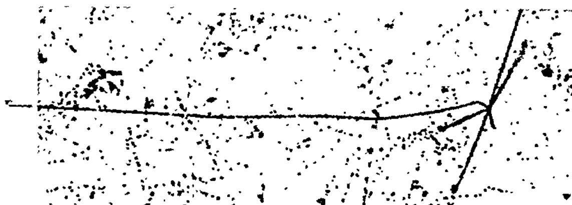
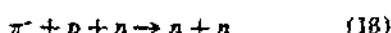


Fig. 3. A meson star produced in photographic emulsion by a π^--meson which enters from left and comes to rest in the emulsion. With five outgoing prongs, this event is unusually conspicuous for a π^- capture star.

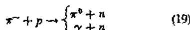
about the nucleus of this atom. It then cascades down through successive orbits until it comes sufficiently close to interact with the atomic nucleus, producing nuclear reactions which release energy and usually disrupt the nucleus. The charged particles emitted in this disruption produce a set of radiating tracks (visible in photographic emulsion) referred to as a star.

The π^- capture stars (Fig. 3) are mainly the result of the strong capture reaction



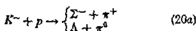
$$\pi^- + p + n \rightarrow n + n \qquad (18)$$

The energy carried by the charged particles which are emitted from these stars as a result of secondary collisions is generally much smaller than the pion rest energy (140 Mev). However, in photographic emulsion, protons with energy exceeding, for example, 30 Mev are observed from 10% of π^--stars. In pure hydrogen, π^- capture reactions involving two nucleons are obviously not possible, and the reactions observed are

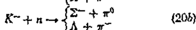


$$\pi^- + p \rightarrow \begin{cases} \pi^0 + n \\ \gamma + n \end{cases} \qquad (19)$$

K^- capture stars result from strong reactions, of which the following are typical examples:



$$K^- + p \rightarrow \begin{cases} \Sigma^- + \pi^+ \\ \Lambda + \pi^0 \end{cases} \qquad (20a)$$



$$K^- + n \rightarrow \begin{cases} \Sigma^- + \pi^0 \\ \Lambda + \pi^- \end{cases} \qquad (20b)$$

Consequently, K^--stars in photographic emulsion are characterized by an appreciable probability for emission of $\Sigma^{\pm}$-hyperons (20%), of $\pi^{\pm}$-mesons (30%), or of both (6%). Λ-hyperfragments are also frequently (4%) emitted from K^- stars.

The μ^- capture stars, on the other hand, are the result of a weak interaction



$$\mu^- + p \rightarrow n + \nu \qquad (21)$$

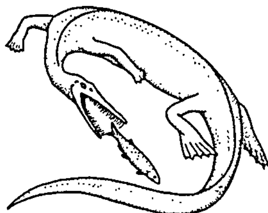
comparable in nature and strength with the β-decay interaction through which certain radioactive nuclei (for example, $_{19}Ca^{41}$) capture K-shell electrons. Since this is a slow capture process, the μ^--meson generally reaches its lowest atomic orbit and spends a considerable fraction of its lifetime there before undergoing decay or nuclear capture. For elements lighter than carbon, the lowest orbit is sufficiently far from the nucleus that the capture process does not compete effectively with the normal decay given by (1). In heavier elements, the formation of capture stars becomes increasingly more probable than decay. On the average, the recoil neutron from the capture process (21) occurring in a nucleus has only 5-Mev kinetic energy, so that the visible energy release in μ^--stars is usually rather small; in fact, the emission of protons from these stars is quite rare. *See* COSMIC RAYS; ELEMENTARY PARTICLE; NUCLEAR STRUCTURE; QUANTUM FIELD THEORY; SCATTERING EXPERIMENTS, NUCLEAR; SYMMETRY LAWS (PHYSICS). [R.H.D.]

Bibliography: M. Gell-Mann and E. P. Rosenbaum, Elementary particles, *Sci. American*, 197:72–88, 1957; R. E. Marshak, Pions, *Sci. American*, 196:84–92, 1957; J. G. Wilson (ed.), *Progress in Elementary Particle and Cosmic Ray Physics*, vol. 3, 1956, vol. 4, 1958.

Mesosauria

An extinct order of aquatic reptiles, also known as Proganosauria, of latest Carboniferous or earliest Permian time, about 240,000,000 years ago. The single known genus, *Mesosaurus*, occurs in Irati bituminous shales and limestones of Brazil and the White Band bituminous shales of South Africa. Though adapted for coastal pelagic life, *Mesosaurus* was probably no more sea-going than the living sea otter, *Enhydra lutris*. Its distribution implies some degree of former connection between South America and Africa. Thus *Mesosaurus* may have ranged around the South Atlantic much as the sea otter ranges around the North Pacific from California through the Aleutian Islands to Siberia. *See* REPTILIA.

The long snout and numerous slender teeth of *Mesosaurus* were well adapted for catching crustaceans and small fish. Its small forelimbs probably served as balancers while the powerful



Restoration of *Mesosaurus*. (*After McGregor*)

paddles and tail propelled it through the water. Its trunk ribs have the swollen form (pachyostosis) characteristic of air-breathing marine animals. Individuals ranged from 2 to 3 ft in length.

Ancestry and relationships of the Mesosauria are uncertain The order may have arisen as an independent offshoot of the Carboniferous "stem reptiles," the captorhinomorph Cotylosauria. It became the first reptilian group to invade marine waters but apparently soon became extinct, leaving no descendants *See* COTYLOSAURIA. [D.BA.]

Mesosphere

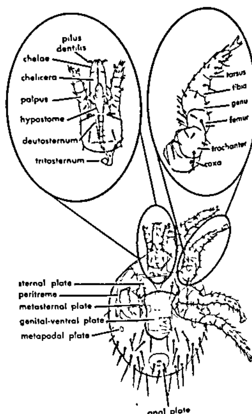
According to international convention, this is the region about 30–80 km, between the stratosphere and the ionosphere, dominated by a marked temperature maximum near 55 km. The temperature at the maximum varies from about 280°K in the tropics to about 250°K in the polar night. The high temperature is caused by absorption of solar radiation by the Hartley bands of ozone. Over most of the earth's surface the region is near radiative equilibrium. Much of the interest in this region of the atmosphere lies in its photochemistry. It is the seat of a number of airglow emissions and most atmospheric ozone is created here. *See* ATMOSPHERE; PHOTOCHEMISTRY. [R.M.C.]

Mesostigmata

A suborder of the Acarina commonly referred to

pair of tubelike peritremes usually runs anteriorly along the idiosoma from the stigmata. The function of these structures is undetermined. A two- or three-tined palpal claw and a tritosternum, which is located ventrally on the idiosoma just behind the gnathosoma, are also diagnostic for this group of mites. The gnathosomal base forms a tube containing the complex mouthparts; and a tectum, a dorsal anterior projection from the gnathosomal base, forms a roof over the mouthparts. The idiosoma is protected by several characteristic sclerotized plates. The dorsum is covered by a large shield which may be entire, divided by a transverse suture into two subequal shields, or divided into a large anterior shield and several smaller posterior shields. Ventrally, the males are well sclerotized and the genital aperture is either within the sternal plate or at its anterior margin. The genital aperture is covered by one or three trap-door-like plates. This is the primary character separating the two major groups, or supercohorts. The Monogynaspida have a single plate, the epigynial. When this has fused with the ventral plate, it is termed the genital-ventral plate. The Trigynaspida have three plates, two latigynials and a mesogynial. When these have been lost or fused, they are replaced by a sternogynial plate, which opens inwardly. The Mesostigmata range from less than 0.2 mm to more than 4 mm in length and are cosmopolitan in distribution.

Development. Mesostigmatid mites generally pass through an egg stage, a 6-legged larval stage, two 8-legged nymphal stages, and an adult stage. Some species give birth to larvae which have passed the egg stage within the body of the mother. Many species are parthenogenetic. Unfertilized eggs produce males while fertilized eggs produce females. In one group of these mites there is a nymphal stage, which serves as a means of dispersal for the species. These nymphs attach themselves to insects

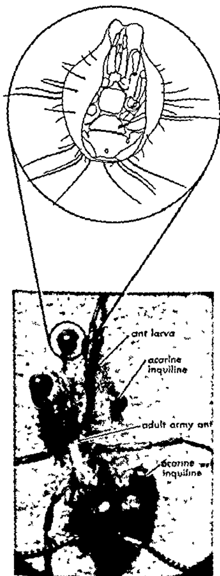


A mesostigmatid mite. (*The Institute of Acarology, University of Maryland*)

and other arthropods by means of an anal pedicel and "hitchhike" to the next suitable habitat.

Interspecific relations. Many of the Mesostigmata are predaceous on other small arthropods or are scavengers and fungus-feeders that make up an important segment of the soil fauna. Some members of the family Phytoseiidae are beneficial to man as predators of the destructive spider mites, Tetranychidae. Other forms are parasitic on insects and other invertebrates.

Terrestrial arthropods are the only invertebrate hosts affected by mesostigmatid mites. As a rule, phoresy is the predominant reason for the association (*see* PHORESY). Complex groups, such as the ants, may have unusual symbiotic relationships with mites, and many bizarre and unusual mite forms have evolved. Apparently the mites contribute nothing to the well-being of the ant colony but are themselves completely dependent upon the ants for their continued survival. A peculiar association exists between certain phalaenid moths and mites of the family Otopheidomenidae. These mites live only in the ears of the moths and are true parasites.



Acarine inquilines on an army ant and larva. (Photograph by Carl Rettenmeyer, The Institute of Acarology, University of Maryland)

Vertebrate hosts are restricted to mammals, birds, and reptiles. The Mesostigmata that parasitize them are of two main sorts; they are either internal parasites of the respiratory passages, or they are ectoparasitic nidicoles, that is, they reproduce in the nests of their hosts. The ectoparasitic mesostigmatid mites may harm their hosts in three ways: they may transmit a disease; they may produce a dermatitis; and their bites result in the loss of blood. In some dermanyssid species, such as the chicken mite and the tropical fowl mite, heavy infestations may cause death of the host by exsanguination, or loss of blood. *Dermanyssus gallinae* (De Geer) is probably the best known parasite of this group and is a pest of poultry, birds, and, less frequently, man throughout the warmer parts of the world It is a vector of avian spirochetosis. Two species found on rodents are of importance as occasional parasites of man: *Ornithonyssus bacoti* (Hirst), the tropical rat mite, and *Allodermanyssus sanguineus* (Hirst), the house mouse mite. The former may cause an irritating dermatitis while the latter is the principal vector of rickettsialpox. The tropical rat mite also serves as an intermediate host of a filarial worm, *Litomosoides carinii* (Travassos), that infests wild rats. In Russia it has been incriminated as a vector of endemic rat typhus. The snake mite, *Ophionyssus natricis* (Gervais), transmits hemorrhagic septicemia from snake to snake. The laelaptine mite, *Echinolaelaps echidninus* (Berlese), and the hemogamasid mite, *Haemogamasus ambulans* (Thorell), are known to harbor and transmit hepatozoons to rats and to squirrels, respectively. *See* articles on specific diseases; *see also* ACARINA.

[J.H.C.; R.W.S.]

Bibliography: A. Treat, Unilaterality in infestations of the moth ear mite, *J. N.Y. Entomol. Soc.*, 68:41–50, 1957.

Mesozoa

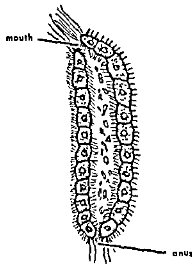
A division of the animal kingdom sometimes ranked as intermediate between the Protozoa and the Metazoa. These animals are unassignable to any of the better known phyla, as usually defined. In the absence of proof concerning their relationships, and in view of the disagreement among zoologists relative to their affinities and even with respect to the facts and interpretation of their structure and life cycle, they are treated simply as a small phylum somewhere between the phyla Protoza and Platyhelminthes. No particular phylogenetic interpretation should be attached to this placement.

The Mesozoa comprise two orders of small, wormlike organisms, the Dicyemida and the Orthonectida. Both are parasitic in marine invertebrates.

The body consists of a single layer of ciliated cells enclosing one or more reproductive cells. These body cells are rather constant in number and arrangement for any given species. The internal cells do not correspond to the entoderm of other animals, as they have no digestive function. The life cycles are complex, involving both sexual and asexual generations (metagenesis).

There has been little agreement regarding the affinities of the Mesozoa. Some zoologists believe them to be related most closely to the Protozoa, others believe them derived from some higher group, usually the flatworms. Other groups, which have also been suggested as a link with the Mesozoa, are the Coelenterata, Echiuroidea, and primitive Aschelminthes. However, the relationship of these groups to the Mesozoa is regarded as less likely by most zoologists.

The Mesozoa have been united into a class called Planuloidea or Moruloidea because they exhibit the same grade of structural organization as a coelenterate planula larva or an embryonic morula.



Longitudinal section of Salinella. (*After Frenzel, 1892*)

A number of other small enigmatic organisms have at times been included in the Mesozoa. Most of these have subsequently been assigned to other phyla. *Salinella*, described by J. Frenzel in 1892, has remained a taxonomic puzzle. It consists of a single layer of cells enclosing a digestive cavity. There is a mouth at one end and an anal opening at the other. Both internal and external surfaces of all cells are ciliated. Asexual reproduction is by transverse fission. Frenzel observed encystment in pairs and described unicellular young resembling ciliates. *Salinella* has not been reported again since Frenzel's discovery. See ANIMAL KINGDOM; CELL CONSTANCY; METAGENESIS; MORULOIDEA.

[B.H.M.]

Bibliography: L. H. Hyman, *The Invertebrates*, vol. 1, 1940; H. W. Stunkard, The Life History and Systematic Relations of the Mesozoa, *Quart. Rev. Biol.*, 29(3):230–244, 1954.

Mesozoic

In geology, the system of rocks younger than the Paleozoic and older than the Cenozoic; also, the geologic era during which those rocks were formed.

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	*Mississippian*	*Pennsylvanian*	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

The Era is now believed to have extended from about 180,000,000 to 70,000,000 years ago. The name Mesozoic was given by early workers in geology to indicate the intermediate nature of its life between the ancient forms in the Paleozoic Era and the more modern life of the Cenozoic Era. The Mesozoic is commonly referred to as the Age of Reptiles inasmuch as these dominated the land, sea, and air. The Era is divided into three systems or periods which are, from oldest to youngest, Triassic, Jurassic, and Cretaceous. See CRETACEOUS; JURASSIC; TRIASSIC.

[W.A.C.]

Messier number

The designation of a nebula or star cluster in Messier's catalog. Charles Messier (1730–1817) was primarily interested in the discovery of comets, and he compiled a list of objects that might be mistaken for comets with the size telescope he used. The first catalog was published in the Royal Academy of Sciences of France in 1771; later lists were published in the *Connaissance des Temps*, the total number of authentic objects being slightly less than 100. Some of the later objects were discovered by Pierre Méchain. Messier's catalog included the Crab nebula M 1, the Andromeda spiral M 31, the Triangulum spiral M 33, the globular star cluster M 31, and the Ring nebula M 57 or NGC 6720. The New General Catalog (NGC) by J. L. E. Dreyer was far more extensive than that of Messier. The NGC numbers are, consequently, generally used for all except a few objects whose Messier number had become well established.

[L.H.A.]

Metabolic disorders

Disorders which involve an alteration in the normal metabolism of carbohydrates, lipids, proteins, water, and nucleic acids. Deviations in normal metabolic processes are evidenced by various syndromes and diseases. See MALNUTRITION.

PROTEIN METABOLISM

Structure and function of proteins. *Proteins* make up the major portion of the body's solid or-

ganic material. They are composed of a series of amino acids bound together in long chains by covalent peptide bonds. The chains are folded in irregular configurations and held to each other by disulfide and hydrogen bonds between the side groups of individual amino acid residues. The great individuality and specificity that is characteristic of cellular proteins results from three major factors. First, each protein contains at least 20 amino acids, each of which may appear in a given protein a varying number of times. Second, the sequence of amino acids may vary. The number of sequence combinations of individual amino acids may be considered to be infinite. Third, the long peptide chains may be folded in a variety of configurations. It is therefore theoretically conceivable that every protein in every individual is different from every other protein. Actually, immunological studies have shown great similarity, although not identity, between proteins, not only in one individual but also between species. This is well illustrated by the protein hormone insulin. The complete structure of this protein, which contains 51 amino acids, has been identified in several species including cows, swine, sheep, horses, and whales. The amino acid sequence in cows and sheep is almost identical; in one small portion the sequence in beef insulin is alanine, serine, valine, while in the sheep it is alanine, glycine, and valine. Most proteins, like other compounds in the body, are complexed with other molecules. They are linked to lipids in the lipoproteins, carbohydrates in the glycoproteins, and nucleic acids in the nucleoproteins. In addition, it is believed that large macromolecular structures incorporating two, three, or four different classes of compounds are present and necessarily interact in an integrated manner for proper functioning of the cell.

Synthesis. There is considerable controversy concerning the mechanism by which proteins are synthesized. It was formerly believed that the union of amino acids into peptides, and a later union of peptides into polypeptides, eventually resulted in a large protein molecule. The persistent inability to demonstrate the presence of intermediate peptides in the cell, and newer knowledge concerning the role of nucleic acids in protein synthesis, have produced a competing theory which is more generally accepted at present. The newer concept suggests that all proteins must arise from single amino acids, perhaps in the following manner. An unknown cellular mechanism transports and degrades plasma proteins intracellularly into a free amino acid pool. Individual amino acids are then activated by an amino acid activating enzyme (a specific enzyme has been isolated for each amino acid) and linked to a molecule of adenosinetriphosphate (ATP). The resultant nucleotide–amino acid molecule is transferred to a ribonucleic acid in the soluble portion of the cytoplasm and then the entire complex is transferred to a particular ribonucleic acid associated with the microsomes, an organelle of the cytoplasm. Here, the sequence of amino acids in the protein being synthesized is presumably determined by the arrangement of the four purine and pyrimidine bases in the nucleic acid. Another enzyme then links the adjacent amino acids by covalent peptide bonding. The long polypeptide chains are released and later are folded in their specific configurations. *See* CYTOLOGY.

Functions of proteins. The functions of proteins are varied. Proteins are present in the major structural component of the cell commonly called the cytoskeleton. This is represented by the cytoplasmic and nuclear membranes of the cell, the particulate organelles such as mitochondria and lysosomes, and a protein network of cisterna and tubules throughout the cell called the endoplasmic reticulum. By definition, enzymes are also proteins. The difference between structural and metabolic enzymatic protein is not distinct in the cell, and there are probably identical proteins which serve both structural and metabolic functions depending on the cellular environment at any one time. Cellular protein is generally considered to be in dynamic equilibrium, that is, a constant state of synthesis and degradation, with its external environment. It is now believed, in addition, that each of the thousand or more proteins in a cell has an individual life span and the rate of synthesis and degradation of a particular protein varies over a wide range depending on the site and function of the protein involved.

Large amounts of protein are produced intracellularly for use elsewhere. The liver cells are the sole site of albumin synthesis, and the serum albumin usually measures approximately 4 g/100 ml of plasma. As one of the plasma proteins, albumin serves a major role in regulating the osmotic capacity of the blood, as a transport system for fatty acids and as a source of amino acids for cellular protein. The other major group of plasma proteins are the α-, β-, and γ-globulins. The α- and particularly the β-globulins are associated with lipids and it is in the latter fraction that the large lipoproteins are found. The γ-globulin fraction contains antibodies. Large quantities of enzymatic protein are produced in the intestinal mucosal cells and in the pancreas for use in the digestion of fats, carbohydrates, and proteins. Other proteins are produced in endocrine glands, particularly the pancreas and pituitary, and function as hormones regulating a multitude of cellular activities. Some of the most important extracellular proteins are antibody proteins which are formed in the reticuloendothelial system. They react with antigens, which are usually proteins but may be proteins combined with polysaccharides or lipids. The antigen-antibody reaction usually results in inactivation of the antigen. *One example of an antigen-antibody reaction* is the neutralization of a bacterial toxin. *See* ANTIBODY; GLOBULIN; OSMOREGULATORY MECHANISMS.

General effects of injury to cell proteins. Cell injury in any form will result in a change in cell protein. The initial insult may be caused by bacteria, viruses, trauma, nutritional deficiency, antimetabolites, or a variety of other agents. R

less of the form of injury, part of the cell's protein usually reacts by becoming denatured, that is, the long peptide chain unfolds. This may result in a lessening of the cell's activities, depending on its innate reserves, its ability to repair the damage, and the particular protein affected. If the injury is slight, the process may be reversible and the cell function will return to normal very quickly. Cloudy swelling is a term applied to cells believed to be in a reversible state of injury. Microscopically, such cells appear swollen and exhibit accentuated cytoplasmic granularity. Some investigators believe that this initial reaction represents swollen mitochondria which have changed their shape from the normal rod to a spherical form. If a more severe injury occurs, the denatured protein may coagulate and protein normally in solution may precipitate. This is considered an irreversible state. Associated with these changes in the protein of the cell ...

becomes irregular, various blebs appear on the surface, and discrete water vacuoles form in the cell. These changes are referred to as hydropic degeneration. The denaturation and disruption of the cell's protein network ...

ac

ad

en ... tissue. The subsequent lack of energy makes it impossible for the cell to maintain its ionic equilibrium. There is a reversal of the normally high potassium/sodium ratio. The cell first swells and then shrinks to a nondescript mass of coagulated protoplasm. The changes in the nuclear proteins may proceed at a different rate from those in the cytoplasmic protein. Several terms have been applied to changes in the nucleus alone: karyolysis, a swelling of the nucleus; karyorrhexis, a fragmentation of the basophilic chromosomal components of the nucleus; and pyknosis, the polymerization and contraction of the nuclear chromosomal components. The mass of denatured coagulated protein, representing the dead cell, is either phagocytized in place by wandering histiocytes or partially dissolved in the extracellular fluid and transported via the lymphatics into the blood stream. Most such particles are removed by the phagocytic activity of the reticuloendothelial cells in the spleen.

Specific effects of injury to cell proteins. While the changes described above may take place very rapidly in experimental situations, various intermediate stages of cell death can be seen in disease states. Lipid or glycogen accumulation, the appearance of hyaline bodies, vacuoles and changes in denatured protein are some of the stages commonly seen. Large lipid accumulations occur in degenerating cells of several organs including the heart, liver, and kidney. In some instances these deposits appear to arise through transfer of lipids from the serum; in other instances they are apparently being newly synthesized. Severely damaged cells lose all their glycogen; some slightly injured cells show large accumulations of glycogen while others merely show vacuolizations which contain water or mucin. In the damage of liver cells associated with nutritional deficiencies, especially when associated with cirrhosis (scarring), the cells may contain large, acidophilic, refractile skeins or discrete bodies near the nucleus which probably represent altered protein. These structures have been called Mallory bodies for diagnostic purposes. In acute hepatitis, another form of liver injury, the entire cytoplasm within the membrane may become acidophilic, homogeneous, and refractile and may eventually be extruded into the liver sinusoids. These bodies in the cytoplasm are known as Councilman's bodies and are useful in differentiating injury by different etiological agents. Whenever cellular or extracellular protein becomes completely and irreversibly denatured and degenerated, it becomes eosinophilic and assumes a homogeneous glassy appearance which is often referred to as hyaline change. It represents an end stage of protein degradation and is found in irreversibly injured cells, inanimate collagen fibrils and other amorphous extracellular protein deposits. Another change, more commonly seen with extracellular than intracellular protein, is the so-called fibrinoid change. Histologically, this change resembles one of the plasma proteins, fibrin, in many ways. It may be injured protein with fibrin deposited on the outer surface, or it may represent a primary change in the protein which imitates the staining characteristics of fibrin. This condition is most commonly seen in diseases concerned with changes in collagen such as rheumatic fever, rheumatoid arthritis and lupus erythematosus.

Protein metabolism disorders. The diseases associated with abnormal protein metabolism have been classified as follows: (1) diseases associated with increased production of proteins, (2) those associated with decreased production of proteins, (3) diseases associated with the production of abnormal proteins, and (4) diseases associated with the excretion of unusual amounts of amino acids.

Most diseases cause changes in protein metabolism and to a greater or lesser extent result in the cellular changes described above. When a cell is injured slightly, the normal reserve components of the cell plus the ability to resynthesize necessary constituents usually result in a quick return to normal cell function and an inability to recognize any changes morphologically. Damage to a few cells in an organ as large as the kidney which contains millions of cells is of little importance in relation to total kidney function and presumably is a common occurrence ...

...ase of protein. ...hen it occurs ... while fewer total proteins, including the α-globulins and the albumins, are synthesized. The diseases usually associated with hyperglobulinemia are multiple myeloma, kala-azar, Hodgkin's disease, lymphogranuloma inguinale, sarcoidosis, liver cirrhosis, and

amyloid. The mechanisms which stimulate the production of proteins, particularly of one type, are not known, but the plasma cell has been implicated. This may be a compensatory phenomenon because of the deficiency of an essential protein whose pathway has been blocked, or it may be an actual loss of control of the mechanisms which normally regulate the amount and type of protein to be synthesized.

Hypoproteinemia. A decreased amount of protein, hypoproteinemia, may be the result of a lack of essential amino acids for protein synthesis, a metabolic block, or other interference with the normal synthesis mechanism. Increased excretion of protein, particularly in chronic renal disease with a loss of albumin in the urine (albuminuria), is another common cause of hypoproteinemia. Kwashiorkor is the best example of hypoproteinemia resulting from dietary deficiency. Since albumin is synthesized in the liver, severe liver inflammation (hepatitis) or scarring (*cirrhosis*) will result in decreased synthesis and hypoalbuminemia or decreased levels of albumin in the blood stream. Since albumin makes up two-thirds of the total protein in serum this will result in hypoproteinemia. Two recently discovered diseases, hypogamma-globulinemia and agamma-globulinemia, may be classified under hypoproteinemia. In these the total serum albumin and globulin are not markedly depressed but the γ-globulins may fall from a normal of 15–20% of the total protein to levels of 0.4%. As previously mentioned, antibodies are normally part of this fraction, and the patients are unable to synthesize immune antibodies against antigens. In addition to this deficiency there is usually an intractable diarrhea in infants, and in adults arthritis and bronchiectasis (dilatation of the terminal bronchioles). Although γ-globulins are produced in the reticuloendothelial systems, the lymph nodes show no disarrangement in the endothelial cells, the main supportive cells of the lymph node. There has been noted, however, a lack of plasma cells. This cell has been most commonly associated with the presence of antibodies.

Formation of abnormal proteins. It is impossible to say whether so-called abnormal proteins represent normal intermediates which are in sufficiently small amounts to remain undetected or whether the proteins really are separate and distinct from any known natural proteins. The primary defect may reside somewhere in the nuclear protein synthesizing system of the cytoplasm, the endoplasmic reticulum, or it may be a result of disturbance in the genic component, desoxyribonucleic acid, of the nucleus.

Myeloma. Multiple myeloma, a neoplastic growth of plasma cells particularly in bone marrow and lymph nodes, is representative of a disease in which an abnormal protein is believed to be produced. The extraordinarily large amounts of γ-globulin present in the serum in this disease may represent only increased, rather than abnormal, synthesis. Many of its physical characteristics are identical with those of normal γ-globulin; however, the Bence-Jones protein found in the urine of multiple myeloma patients differs markedly from any protein found normally in the serum or urine. Its molecular weight is approximately 40,000, one-fourth that of normal γ-globulin (160,000). The use of radioactive glycine has shown that this protein in the urine becomes labeled with radioactive carbon before the plasma globulins do, indicating that it is not merely a degradation product of the plasma protein but is either an entirely new protein or a protein formed as an intermediate on the normal pathway of the larger plasma proteins. *See* NEOPLASIA.

Hemoglobins. Another group of abnormal proteins have been associated with the hemoglobins. Hemoglobin is a protein with a molecular weight of approximately 68,000 containing four porphyrin groups linked to an iron molecule. It functions in the transport of oxygen and carbon dioxide in the blood. Several different hemoglobins in red blood cells in mammals have been described, one of which produces the disease called sickle-cell anemia. The nine different forms of hemoglobin discovered thus far are distinguished by changes in physical characteristics and are named A for adult, F for fetal, S for sickle cell, C, D, E, G, H, and I. The differences between these hemoglobins apparently reside in the arrangement of the amino acids and the folding of polypeptide chains. The sickle-cell hemoglobin, for instance, has a replacement of the glutamyl amino acid residue by a valyl residue. *See* HEMOGLOBIN; PORPHYRIN.

Amyloid. This material takes on a homogeneous, eosinophilic color when stained with hemotoxylin and eosin. It was originally thought to be starch and later chrondroitin sulfuric acid. In contrast to the two abnormal proteins described above (Bence-Jones proteins of multiple myeloma and the abnormal globin in sickle-cell anemia), amyloid probably represents not a defect in synthesis but an abnormal degradation product of protein. It has also been suggested that it is an abnormal combination of proteins, perhaps an antigen-antibody complex. Its exact composition is unknown but it is believed to contain carbohydrates and to be a glycoprotein. The theory that it is an antigen-antibody complex, formed in response to an autogenous stimulation from dying cells, stems from its being commonly found in diseases where degenerating cells are present. Deposits of amyloid were formerly classified as primary or secondary, depending on the sites of deposits and whether there was a chronic inflammatory reaction present elsewhere in the body. Thus secondary amyloid was usually associated with deposits of amyloid material in the spleen, liver, and kidney, and primary amyloid was associated with deposits in the tongue, mesenchymal tissues, and upper respiratory organs. It is presently thought that there is considerable overlapping between the types, not only in the sites of deposit but also in their being secondary t chronic disease. In many organs deposits loid may lie free in the extracellular s idney the material is usually de

extracellular space between the basement membrane and the endothelial lining of the glomerulous; in the liver it lies between the sinus endothelium and the hepatic cell.

Amino aciduria. This is a group of disorders in which there appears to be an increase in the amount of amino acids excreted in the urine. They are due to abnormal protein metabolism and result either from an overflow mechanism where the concentration of amino acids in the serum surpasses the renal threshold of the glomerular membrane, or from defective absorption of amino acids in the renal tubules. The overflow amino acidurias include phenylketonuria, alcaptonuria, and liver disease.

Phenylketonuria. This is a disease in which large amounts of phenylpyruvic acid, a degradation product of the amino acid phenylalanine, is found in the urine of patients who may also be mentally deficient. The biochemical defect in this disease is an absence, presumably hereditary, of the oxidase enzyme which converts phenylalanine to tyrosine. The accumulation of phenylalanine eventually results in accentuation of alternate pathways of phenylalanine degradation leading to phenylpyruvic acid, phenylactic acid, and phenylacetate. The latter products are present in such excessive quantities in the serum that they are excreted in the urine.

Alkaptonuria. In this similar disease, there is again failure of complete breakdown of phenylalanine and tyrosine; an intermediate product, homogentisic acid, accumulates in the tissues and serum and is excreted in the urine. It is likely that the enzyme homogentisase which converts homogentisic acid to fumaryl acetoacetic acid is either not formed in alkaptonuria or is in some way inhibited. The deposits of homogentisic acid in the tissues, especially in the cartilages, tendons, ligaments, and sclerae of the eyes, give a bluish discoloration known as ochronosis.

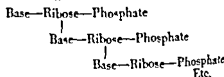
Liver and kidneys. Since the liver plays such a major role in the deamination of amino acids, gross necrosis or advanced cirrhosis may lead to increased levels of amino acids in the blood and subsequent excretion in the urine.

Other diseases having amino aciduria are believed to be the result of defective kidney function. The defect seems to lie in the inability of the renal tubules to reabsorb amino acids from the glomerular filtrate. Thus, in cystinuria, the failure to reabsorb cystine, lysine, arginine, and ornithine results in their appearance in the urine. A similar situation obtains in Wilson's disease, a degeneration involving the liver and brain, although the amino aciduria is highly variable in this disease. Amino aciduria may also occur in Fanconi's syndrome, galactosemia, scurvey, rickets, and toxic poisoning with lead, cresol, and benzene.

NUCLEIC ACID METABOLISM

Nucleic acids. Originally defined as the acid constituents of the nucleus, nucleic acids are now known to be present throughout the cell. They are divided into two major categories. Deoxyribonucleic acid (DNA) is found principally in the chromosomes of the nucleus of mammalian cells and is closely identified, if not synonymous, with the genic material. Ribonucleic acid (RNA) is found in both the nucleus, especially in the nucleolus, and the cytoplasm, and it is believed to be closely associated with the process of protein synthesis. Within these two broad categories there may be several subclassifications according to molecular weight and other physical characteristics.

Structure. Each nucleic acid molecule contains purines and pyrimidines bound to a sugar, either ribose or deoxyribose, and linked together by phosphate bonds. The purines and pyrimidines are referred to as bases; when linked to the sugar the complex is called a riboside; and when the phosphate is added, a nucleotide. Hundreds of nucleotides make up a molecule of nucleic acid. Both DNA and RNA contain the purines adenine and guanine and the pyrimidine cytosine. For the second pyrimidine, RNA contains uracil while DNA contains thymine. The structure of RNA is usually depicted as follows:



Diseases. Since the nucleic acids are closely associated with the chromosomes and the process of cell division as well as with protein synthesis, they are one of the most important of cell constituents. They have been shown to be very susceptible to injury by various agents and interruption of division is one of the earliest signs of cell damage. A disturbance in nucleic acid metabolism is of extreme importance to many other aspects of cell metabolism; however, at present there are only two diseases, gout and lupus erythematosus, which appear to be associated with disturbances in nucleic acid metabolism. But the whole field of virus infection of mammalian cells is undoubtedly largely concerned with a disturbance in nucleic acid or nucleoprotein metabolism.

Gout. Gout is a form of arthritis, associated with extreme pain, especially around the smaller joints. There are usually accumulations of uric acid in the blood (hyperuricemia) and deposits of uric acid (tophi) in the soft tissues surrounding the joints, the cartilage of the ear, the diaphysis of bones, the kidneys, and the heart valves. The pathognomonic tophus of gout is readily identified as a collection of sodium urate crystals surrounded by an inflammatory response. The hyperuricemia must result from increased production or decreased excretion of uric acid or a diminished conversion of uric acid to urea. Increased production of uric acid, a degradation product of purines, has been determined in some cases.

Lupus erythematosus. Lupus erythematosus may be classified under abnormalities of either protein or nucleic acid, since it undoubtedly involves a disturbance in nucleoprotein metabolism. The etiology is unknown but the patients are extremely photosensitive and have hyperglobulinemia. There is [illegible]

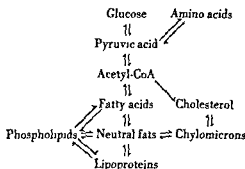
lesion shows a granular, friable exudate of fibrin with increased amounts of amorphous ground substance. In some instances eosinophilic necrotic material resembling fibrin is seen coating the collagen fibrils on their free surfaces. Within the fibrinoid material are dark smudges resembling hematoxylin bodies. There is now a reliable diagnostic test, known as the lupus erythematosus phenomenon, available for this disease. When serum from a lupus erythematosus patient is mixed with white blood cells from a control individual, the neutrophilic white blood cells develop large basophilic inclusions which are believed to be nucleoprotein. The significance of this particular nucleoprotein, its origin, or its function is unknown but it does represent an abnormality in nucleic acid metabolism.

Virus infection. Little is known concerning alterations in cell metabolism when a virus enters a cell and much of our existing knowledge comes from the study, not of mammalian viruses, but of viruses called bacteriophages which infect bacterial cells. Viruses are believed to contain few if any enzymes and are parasites since they are dependent on the host cells mechanism to synthesize the metabolites necessary for their replication. Although the viruses have small amounts of protein, the nucleic acid component is believed to be the material responsible for entering the cell and producing infection. A virus may stimulate cellular activity, including growth, may have no apparent effect on cells, or may cause cytolysis and death of cells. The latter phenomenon usually occurs after the virus has multiplied many times and is being released from the cytoplasm. Tissue culture studies have shown an increase in both RNA and DNA as well as protein when the cell is infected with certain viruses. The continued study of viral infections should show that they produce marked effects on nucleic acid metabolism. *See* BACTERIOPHAGE.

LIPID METABOLISM

Structure of lipids. The basic unit of lipids is the fatty acid which may contain as few as three carbons in propionic acid or as many as 18 carbons in stearic [illegible]
in [illegible]
(CH [illegible]
and [illegible]
$(CH_2)_7COOH$. Arachadonic, linolenic, γ-linolenic, and linoleic acid are considered to be essential fatty acids in that the organism cannot synthesize them. Fats as they exist in the body are generally classified as free fatty acids, neutral fats, chylomicrons, phospholipids, or cholesterol, although there is considerable overlapping of the subdivisions. Free fatty acids probably do not exist alone but are usually bound to proteins as lipoproteins in the plasma or in the cell. Much of the fat is linked to a serum or cellular protein in the neutral fats, which consist of three molecules of fatty acids bonded to three molecules of glycerol by an ester linkage. Large molecules of neutral fat surrounded by a thin layer of protein form microscopically visible globules called chylomicrons. Lecithin is a phosphatidic acid linked to choline and is a representative of the phospholipids. The steroid cholesterol, although not a true fat, is usually included in the lipid group.

Fatty acids may arise from degradation of fats in the intestinal tract or be synthesized from acetate, a degradation product of glucose. When acetate is linked to coenzyme A, a product acetyl-CA is formed, which is the basic building unit for all fats.



The interrelationships between the various lipid compounds are shown in the above diagram. The cell may degrade one class of lipids if synthesis of another group becomes necessary. Lipids are probably all transported with proteins, either loosely or strongly bound. It has been suggested that a lipoproteinase released by the cells (clearing factor), helped by a coenzyme, heparin, breaks not only the lipoprotein bonding but also the fatty acid glyceral bond of the neutral fat. By this mechanism fatty acids may be transferred from their lipoprotein complex in the serum to the lipoprotein complexes in the cell. Other mechanisms such as pinocytosis, active transport, or diffusion may play a role in the transport of lipid across the cellular membrane. Once in the cell, fats are presumably degraded, oxidized for energy, or used for synthesis of other lipids, carbohydrates, or proteins by alternate metabolic pathways.

Function of lipids. The function of fat is both structural and metabolic. Most of the cell membranes are considered to be lipoprotein complexes. The degradation and oxidation of fats play a role secondary only to carbohydrates in the formation of adenosinetriphosphate (ATP). Although cholesterol appears in most cells and is present in large

amounts in the serum, its function remains unknown.

Obesity, hyperlipemia, lipid tumors. Obesity in general may be attributed to an excess intake of calories and only rarely may glandular deficiencies such as hypopituitarism and hypothyroidism be implicated as its cause. Increased caloric intake results in an increase in the size and number of fat cells distributed throughout normal lipid deposits and many organs of the body.

[illegible] primary type of lipemia is caused by a genic defect. There is marked delay in the clearing of ingested fat so that a milky serum with elevation of all the lipid fractions, especially of the neutral fat, occurs. Severe sclerosis, and occasionally xanthomatosis, is common in this disease in early adulthood. Another primary hyperlipemia, called idiopathic hypercholesteremia, has also been described. Sclerosis of all the vessels is marked and the patients have a high incidence of myocardial infarct.

There are small tumors composed of fat cells called lipomas which are commonly found in the subcutaneous tissue and which occasionally become [illegible] in sufficient concentration to be seen in normal cells, except for fat cells, under the microscope. In many disease states, fat appears in cells in excessive amounts and becomes readily visible under the microscope when the cells are examined in the living unstained state. During the fixation and staining of microscopic sections, organic solvents are used which remove the lipids of the cell. Thus in hemotoxylin and eosin stained sections, the space left after the fat has been dissolved presents a well-circumscribed clear vacuolar space. In frozen sections in which the lipid is not removed or in tissues which are not subjected to organic solvents, fat will stain avidly with the Sudan stains. The accumulations of fat in cells have been variously named fat accumulation, fat phanerosis, and fat degeneration. Fat in cells may arise either endogenously from new synthesis or exogenously from the environment surrounding the cell. The origin cannot be determined by examining the tissue alone.

Most cells are capable of synthesizing lipid and some cells apparently produce excessive amounts under abnormal conditions. This excess lipid may remain in the cell or be released to the extracellular fluid and thence to the blood stream. Elevated blood lipids and cholesterol may be relatively independent of the diet. The excess lipids present when the dietary lipids are severely restricted must be produced by cells, although not necessarily by the cells in which they ultimately appear. If cells do not manufacture and store lipid within themselves they must obtain the lipid from extracellular sources, presumably the lipoproteins of the serum.

Lipids and degenerate cells. There are four common major diseases which are associated with accumulations of lipid in and around damaged cells. These are, in order of importance, arteriosclerosis, myocardial infarct, cirrhosis of the liver, and nephrosis.

Arteriosclerosis represents a granuloma of proliferated fibroblasts, chronic inflammatory cells, hyalinized collagen, and large lipid deposits which may ultimately ulcerate or calcify. There has been considerable controversy as to whether the collagen first degenerates and the accumulation of lipid is secondary to the denatured collagen or whether the lipid accumulation is primary and the fibroblastic proliferation secondary to the presence of lipid. Most of the lipid in arteriosclerotic plaques appears extracellularly. This fact, in addition to the extraordinarily large amounts of lipid which may be present, lends favor to the hypothesis that lipid is being brought in from the serum rather than being newly synthesized in the connective tissue cells of the aorta. However, the fact that the plasma contributes lipid probably means increased synthesis elsewhere in the body. It seems likely too that a change occurs in the collagen or extracellular ground substance which is more conducive to lipid accumulation and precipitation. The primary change remains cellular in nature and fluctuating serum lipid levels reflect the altered cellular metabolism.

Myocardial infarct results from thrombosis superimposed on severe arteriosclerosis in the coronary vessels. An accumulation of lipid appears intracellularly in the myocardial cells deprived of oxygen, in addition to the lipid in the vessel wall mentioned above. The origin of this intracellular lipid is not known. Since the myocardial cells are not in direct contact with serum and since the large accumulation of lipid does not appear extracellularly, it is conceivable that this lipid arises newly as a result of synthesis from acetate. However, cells in which energy metabolism is altered experimentally with metabolic inhibitors show large accumulations of lipid from extracellular lipoprotein and ischemic damaged cells probably also incorporate exogenous lipid.

In cirrhosis of the liver, scarring is secondary to liver cell destruction and is commonly associated with nutritional deficiencies. Lipid appears as small and large droplets throughout the damaged liver cells. Again as in arteriosclerosis, there is close physical association with the plasma lipids (the liver cells border the sinusoids containing plasma) which favors an exogenous source of lipid. However, this lipid accumulation is known to be associated with protein deficiencies and substrates which normally go into protein metabolism may be diverted into lipid synthesis instead. Tissue culture cells in protein deficient media accumulate lipid as a result of synthesis from acetate. The process of lipid accumulation in the liver is reversible when it is produced with protein deficiencies, but not when it is produced with toxic chemicals. Identical

results are seen experimentally in tissue culture cells.

The accumulation of lipid in the kidney tubules in lipoid nephrosis, membranous glomerulonephritis, again is representative of a situation in which injured cells are close to the plasma lipids. The arguments for and against exogenous versus endogenous accumulation in the kidney are very similar to those advanced for the toxic damage in the liver.

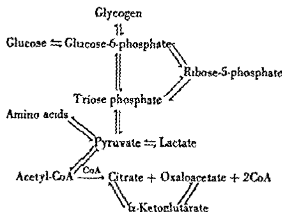
Lipid storage disease. There is another group of diseases in which lipid accumulates because of a disturbance in lipid metabolism, not dependent as far as we know on external stimuli. These are the so-called lipid storage diseases which are in some ways very similar to the glycogen storage disease. In all of these diseases, a large accumulation of lipids appears in many cells, but particularly in the reticuloendothelial cells of the lymph nodes, liver, spleen, and bone marrow. The lipid seems to be distributed throughout the cell and is not well localized. It should be reiterated that in Niemann-Pick and Gaucher's diseases the primary defect is probably in nucleoprotein metabolism which has erred by producing deficient or abnormal enzymatic protein to regulate lipid metabolism and has caused great masses of lipid to be stored in the cells. In Niemann-Pick disease, a diaminophosphatide, sphingomyelin, is found to be the principle lipid present. In Gaucher's disease an abnormal glycolipide, cerebroside, is the substance stored.

CARBOHYDRATE METABOLISM

Structure and function of carbohydrates. Carbohydrates form a large part of all mammalian diets and play a major role in supplying the necessary caloric requirements for maintenance and growth of the individual organism. The best known carbohydrate is glucose ($C_6H_{12}O_6$), a monosaccharide which polymerizes to form two important polysaccharides, starch, found in plant tissues, and glycogen, found in animal tissues. In mammals, glycolipoproteins, polysaccharides bound to proteins and lipids, are important as macromolecular structural units of membranes and particulate organelles of the cell. Polysaccharide-protein complexes are capable of acting as antigens in the stimulation of antibody production; polysaccharides are also a major component of the extracellular ground substance in the body.

Although both fat and protein may serve as a source of energy for cellular metabolism, the simple sugar glucose is the major contributor to the energy pool. Each glucose molecule has three major alternative pathways to follow after entering the cell (see diagram).

Glucose may polymerize to form glycogen for use as a reserve store of energy, be converted to ribose for use in synthesis of nucleic acid, or be degraded to pyruvic acid. During periods of increased work or oxygen lack, the pyruvate may be converted to lactic acid. Normally, the pyruvate is converted to acetyl-coenzyme A which is used in the synthesis



of fatty acids or, after linking with oxalacetic acid, is oxidized to carbon dioxide and water. One of the intermediates in the oxidation of carbohydrates is α-ketoglutaric acid which is interconvertible with glutamic acid, an *amino* acid. Thus the intermediary metabolism and pathway of carbohydrates, fats, and proteins are closely interrelated.

Carbohydrate deficiency. Small molecules in the cell, including sugars such as glucose, fructose, and ribose, cannot be seen with the magnification and resolution powers of the light microscope although they are readily detected by chemical analyses and chromatographic techniques. The sugar that is seen most easily in the cell is glycogen. This metabolite is particularly prominent in the liver and muscle cells and may appear in a diffuse lace network pattern or as discrete vacuoles, most clearly demonstrated with Best's carmine stain. Other polysaccharides, both intracellular and extracellular can be seen with special stains. A few of the enzymes concerned with carbohydrate metabolism, notably succinic dehydrogenase, have been identified intracellularly by histochemical methods, but these are, of course, proteins.

A lack of glucose quickly results in a lack of adenosinetriphosphate (ATP) production and a lack of energy for the cells' multitudinous reactions. The cell is unable to maintain its necessary ion concentrations, synthesis ceases, and degenerative changes quickly appear in the protein and nucleic acid structural components. There are undoubtedly many undiagnosed states in which relative deficiencies of energy exist but which are not severe enough to produce clinically recognizable symptoms. Individual cells may die, parts of some organs may cease functioning, and adaptive alternative metabolic pathways may be utilized in protein and lipid metabolism when energy requirements are not fulfilled by carbohydrates. Most of the diseases ascribed to alterations in carbohydrate metabolism are in reality abnormalities in protein or nucleic acid metabolism. The primary defects are deficient or abnormal enzymes, and the resulting disturbances in carbohydrate metabolism are the effect of this deficiency rather than its cause.

Abnormal carbohydrate metabolism. There are five pathological states which are reasonably well

elucidated in terms of deficient mechanisms or in terms of the specific carbohydrate involved: diabetes mellitus, glycogen storage disease (von Gierke's disease), Hurler-Pfaundler's disease, galactosemia, and malignant neoplasm. The end effects of these diseases have been known for many years; the causes still remain obscure.

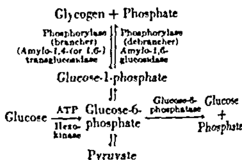
Diabetes mellitus. Diabetes mellitus is a disease of carbohydrate metabolism in which there is a deficiency of the protein hormone insulin. This deficiency results in an inability of the cells to utilize glucose necessary for energy requirements and a hyperglycemia, increased amounts of glucose outside the cell in the blood stream. To obtain the energy for maintaining the body's activities, the cells oxidize fat with resultant ketonemia, ketone bodies in the blood, and ketonuria, ketone bodies in the urine.

Since insulin is produced by specialized β-cells in the islets of Langerhans, which are scattered throughout the pancreas, logically these cells should be damaged or absent in diabetic states. In some cases of diabetes microscopic examination of the pancreas shows a diminution in the size and number of the islet cells, hydropic vacuolation of the cells, hyaline degeneration (a type of protein denaturation), and replacement fibrosis accompanied by lymphocytic infiltration. However, in 25% of the cases of severe diabetes, the islets appear completely normal. One can only conclude that a functional alteration exists which is not manifested structurally or cannot be demonstrated with existing knowledge of the cell. In a small number of cases, glycogen is found in the convoluted tubules of the kidney and is thought to represent an increased absorption of glucose from the glomerular filtrate. *See* PANCREAS.

Glycogen storage disease. Small quantities of glycogen, a polymerization product of glucose, may be found in many cells, where they act as a storage supply for energy when free glucose is unavailable. In glycogen storage disease extraordinarily large amounts of glycogen are deposited in essential organs such as the heart, liver, kidney, and skeletal muscle. The symptoms of the disease arise when the accumulation of glycogen interferes with the normal cells' metabolism. When this occurs in the heart, the muscle fibers appear split by large amounts of glycogen, muscle contractility is impaired, and heart failure and death may eventually result.

The disease has been classified into four general types although there is considerable crossing over of the classification. The accompanying diagram illustrates the pathways of glucose in glycogen synthesis and will make identification of the four major types easier.

In classical von Gierke's disease, the major organ involved is the liver. It may contain as much as 16% glycogen wet weight, but the glycogen has a normal structure. The enzyme glucose-6-phosphatase is deficient so that glucose-6-phosphate is

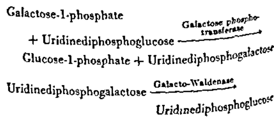


Pathways of glucose in glycogen synthesis

not broken down to free glucose in the blood and an increased synthesis of glycogen results. The liver is also the site of the second form of the disease, but in this instance the glycogen has an abnormal structure with either short or long outer chains in the frozen molecule. These abnormal structures result from a deficiency of the debranching enzyme, amylo-1,6-glucosidase, and the branching enzyme, amylo-1,4-(or 1-6-)transglucosidase, respectively. The third type of the disease is characterized by generalized glycogen storage throughout several organs of the body. The heart, in particular, shows a lacework appearance of the myocardial fibers. The glycogen is apparently normal and the presumed enzyme defect has not been elucidated. The fourth and rarest type is glycogen disease of skeletal muscle which results in progressive muscle weakness. In a few cases an excess of a low-molecular-weight component of the glycogen has been found.

Hurler-Pfaundler's disease. There is excess storage of a mucopolysaccharide in the spleen, liver, and other tissues in this disease, often called gargoylism. The mucopolysaccharide consists of a family of oligosaccharides, each of which is composed exclusively of D-glucosamine and D-glucuronic acid united in glycosidic linkage. The symptoms, as in glycogen storage disease, result from interference with the cells' function. The enzymatic deficiency in this disease is still unknown.

Galactosemia. This is a disease of the newborn resulting from an inherited deficiency of the enzyme concerned with the conversion of galactose to glucose phosphogalactose transferase. The principal food of newborns is milk which contains large quantities of lactose. Lactose is usually broken to galactose and glucose, and the glucose is quickly utilized. The galactose, however, must first be converted to glucose. If the galactose is not converted it accumulates in cells and interferes with their normal functions.



The deficiency of phosphogalactose transference cannot be directly treated at present. If milk is continued in the diet, severe liver damage, cataracts, and mental retardation may result. However, the child with galactosemia can usually be cured if lactose is immediately eliminated from the diet.

Malignant neoplasms. For many years it has been recognized that malignant neoplasms had abnormalities in carbohydrate metabolism. Malignant tissues, to maintain their rapid state of growth and division, require a considerable amount of glucose as an energy source. There are, however, some tumors in which respiration and aerobic oxidation of glucose are inhibited by excessive amounts of glucose, the Crabtree effect. Normal tissues produce lactic acid from glucose anaerobically; only cancer tissue shows the unusual ability to produce lactic acid from glucose in the presence of oxygen, the Pasteur effect. It has been suggested that the Pasteur phenomenon results from a deficiency of the respiratory enzymes' oxidizing pyruvate to carbon dioxide and water and an acceleration of conversion of glucose to lactic acid by the glycolytic enzymes. No adequate proof for a respiratory deficiency has as yet been presented. Usually the cancer cell is easily distinguished from nonneoplastic cells by its morphology involving characteristics of the chromosomal pattern, size of the nucleus, and irregularity of cell outline. Unfortunately there appears to be no correlation between the abnormal carbohydrate metabolism and the morphological structure of the cell. The carbohydrate abnormalities are believed by many investigators to be secondary to metabolic abnormalities of nucleic acid and protein, including the factors involved in the regulation of the cells' division process. [D.W.K.]

Bibliography: D. H. Brown, Tissue storage of mucopolysaccharides in Hürler-Pfaundler's disease, *Proc. Natl. Acad. Sci.*, 43:783–790, 1957; G. R. Cameron, *New Pathways in Cellular Pathology*, 1956; J. S. Fruton and S. Simmonds, *General Biochemistry*, 2d ed., 1958; J. P. Greenstein, *Biochemistry of Cancer*, 2d ed., 1954; H. M. Kalckar, E. P. Anderson, and K. J. Isselbacher, Galactosemia, a congenital defect in a nucleotide transferase: a preliminary report, *Proc. Natl. Acad. Sci.*, 42:49–51, 1956; R. H. S. Thompson and E. J. King, *Biochemical Disorders in Human Disease*, 1957.

Metabolism

All the physical and chemical processes by which living, organized substance is produced and maintained and the transformations by which energy is made available for use by the organism.

In defining metabolism, it is customary to distinguish between energy metabolism and intermediary metabolism, although the two are, in fact, inseparable. Energy metabolism is primarily concerned with over-all heat production in the organism while intermediary metabolism deals with chemical reactions within cells and tissues. In general, the term metabolism is interpreted to mean intermediary metabolism.

Metabolism thus includes all biochemical processes within cells and tissues which are concerned with their building up, breaking down, and functioning. The synthesis and maintenance of tissue structure generally involves the union of smaller into larger molecules. This part of metabolism, the building of tissues, is termed anabolism. The process of breaking down tissue, of splitting larger protoplasmic molecules into smaller ones, is termed catabolism. Growth occurs when anabolism exceeds catabolism. Weight loss results if, on the other hand, catabolism proceeds more rapidly than anabolism, as in periods of starvation, serious injury, or disease. When the two processes are balanced, tissue mass remains the same.

Various components of the tissues are continually undergoing degradation in catabolism and resynthesis in anabolism. Some of the chemical reactions involved in these metabolic processes are exergonic, that is, they are accompanied by liberation of energy, whereas others are endergonic, that is, require energy. The endergonic reaction is the so-called energy of activation. After energy has been introduced into an endergonic process to move substances from a metastable state or state of intermediate stability, an exergonic reaction may occur, which, in addition to replacing the energy of activation, releases additional energy. Within cells, for example, there exists a mechanism whereby free energy, available from oxidative reactions, may be used to spark endergonic processes (*see* BIOLOGICAL OXIDATION). This is accomplished by trapping this energy in the formation of a special class of phosphate compounds of high energy content, the high-energy phosphate bonds ($\sim$P); *see* ACETYL PHOSPHATE; ADENOSINETRIPHOSPHATE (ATP). In some oxidative reactions, however, it seems that the released energy cannot be used for endergonic purposes. The aerobic dehydrogenases of cytoplasm act as catalysts in the oxidation of a number of substrates; these processes serve in the synthesis of needed protoplasmic substances or in the conversion of materials into excretable form, but apparently they cannot be utilized in endergonic reactions.

Metabolism is largely concerned, particularly in the case of synthetic or anabolic phenomena, with the need for providing energy-producing mechanisms for the construction of new tissues and their constituents. The energy is often obtained from simultaneously occurring degradative, or catabolic, reactions. Exergonic transformations are utilized for synthesis of all structures, for maintenance of body temperature, and to provide energy for specific functioning of various tissues, as in muscle contraction, nerve impulse conduction, glandular secretion, absorption, light production, and excretory processes. To replace the energy thus utilized, a

regular provision of new external sources of energy in the form of various foodstuffs must be supplied.

The metabolism of the three major foodstuffs, carbohydrate, fats, and proteins, is intimately interrelated, so any clear-cut division of the three is arbitrary and inaccurate. Thus, the metabolism of protoplasm is concerned with all three of these foodstuffs. However, each function of an organism has its own particular metabolic needs. For example, the chemical reactions of major importance in muscle contraction are those which yield energy for transmission to and operation of the contractile elements in muscles. The high-energy phosphate bonds (∼P), formed during metabolic degradation and transformation of carbohydrate, fatty acids, and amino acids, transfer their energy to the muscle fibrils and so cause contraction.

The metabolism of carbohydrates, fats, and proteins has many points of intersection; thus certain pathways of metabolism are shared in common by fragments of these different classes of foodstuffs. For example, breakdown products of carbohydrates,

ble, chemical energy, in the form of ∼P, must be supplied. The major sources of this energy in respiring tissue are the citric acid cycle and the cytochrome systems, a series of oxidative and degradative reactions. It is through these metabolic systems that most of the ∼P is produced. Carbon dioxide and water are by-products of these reactions.

Many of the metabolic processes of the protoplasm of both plant and animal cells follow common pathways; carbohydrate metabolism in plants is similar in many details to carbohydrate metabolism in animals. Therefore, the study of metabolism

CYCLE; LIPID METABOLISM; PLANT METABOLISM; PROTEIN METABOLISM. [M.B.MC.]

Metabolism in ruminants

...the fermentation in the rumen is so active that all saccharide material is converted to short-chain fatty acids and little or no glucose survives for absorption in the gut (see illustration). This imposes an unusual pattern on the intermediary metabolism of ruminants. Calculations show that 90% of the energy metabolism in these animals is met by combustion of circulating acetate rather than of glucose as in other mammals. The absence of glucose in the gut is also reflected in the low blood-glucose level in the cow (40–50 mg% compared with 100 mg% in man), and in the relatively high short-chain fatty acid content (12 mg% as acetic acid). See CARBOHYDRATE METABOLISM; LIPID METABOLISM; METABOLISM.

Cellular metabolism of acetate. The tissues of nonruminants can also use acetate, but normally it appears only as a fleeting intermediate coupled with coenzyme A in the catabolism of glucose and fats. Experiments show that ruminant tissues can utilize acetate for energy metabolism more readily than can nonruminant tissues. It is unlikely that the enzyme systems of ruminants differ intrinsically from those of other closely related mammals, but rather that the ruminant cells adapt to the pattern of nutrients with which they are presented. This is borne out by the lesser utilization of acetate in the young calf before the rumen microflora (and the large acetate production) develop. The adaptation is probably concerned with the coupling of coenzyme A to "raw" acetate, after which the metabolism of the acetate could proceed on normal lines.

Glucose synthesis. In spite of this efficient utilization of acetate in ruminants, glucose is still a necessary metabolite for some functions. The carbohydrate stored in the liver as glycogen must still be synthesized from glucose units and there is evidence that the ruminant's nervous tissue is not so independent of glucose as are the other tissues. The greatest demand for glucose, however, is in the synthesis of lactose (milk sugar) in the lactating animal. It has been calculated that the "pool" of glucose in the body fluids of a high-producing milk cow is turned over about every hour for lactose synthesis alone. Because so little glucose is available from the gut, it follows that it must be synthesized from other materials.

Work with carbon-14-labeled metabolites has shown that all of the short-chain fatty acids are precursors of the glucose and galactose moieties of lactose in ruminant milk. The actual proportion of acetate converted to glucose is small, but when the volume of acetate production is considered, it is undoubtedly an important source of lactose. Glucose synthesis from the lesser amounts of propionate and butyrate is even more efficient, giving evidence that this process occurs by pathways not involving prior degradation to acetate. Isolated tissue studies show that butyrate and propionate are actively metabolized by the rumen mucosa where they are absorbed.

The pattern of labeling in glucose and certain amino acids after the injection of carboxyl- and methyl-labeled acetate suggests that the synthesis proceeds by way of the tricarboxylic acid cycle and the glycolytic pathways. These are usually thought of as catabolic pathways only, but in the cow, and to a lesser extent in other animals, a reversal of these pathways is also important in synthesis. Even the carbon dioxide present in the tissue fluids is fixed in some of the synthetic processes, although of course the energy for fixation comes from the oxidation of other metabolites and so cannot be

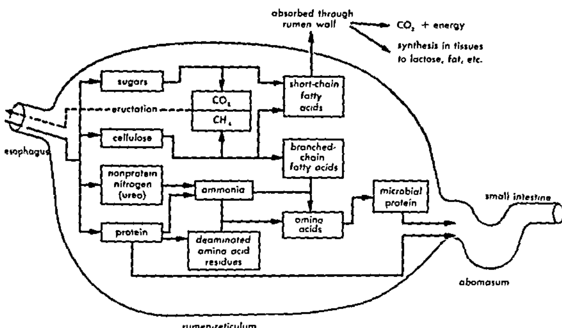


Diagram of rumen fermentation. (From T. A. Rogers, The metabolism of ruminants, Sci. American, 198(2). 34–38, 1958)

compared with the photosynthetic reduction of carbon dioxide in plants. *See* KREBS CYCLE.

Fat synthesis. In milk-fat synthesis, the cow uses acetate and butyrate units to construct the long-chain fatty acids and also uses acetate, propionate, and butyrate as the precursors of the glycerol moiety of the fat. In bovine mammary gland slices, fat synthesis can proceed with acetate as the sole substrate, which shows that it is not only the source of the two-carbon units but the source of energy and glycerol as well. On the other hand, in nonruminant mammary gland slices (guinea pig), fat synthesis requires the presence of glucose, not only for energy but also for the glycerol moiety. *See* LACTATION; MAMMARY GLAND.

When radioactive glucose is injected into a cow, the milk sugar is highly radioactive, the fat only moderately so. Furthermore, most of the fat radioactivity is in the glycerol. The labeling of expired carbon dioxide is only a fraction of that following a comparable injection of labeled acetate. These results point out the peculiar intermediary metabolism of the ruminant. When nonruminants synthesize fat in the presence of labeled glucose, the fat-fatty acids are highly radioactive because they are composed of acetate units derived from the degradation of glucose. The cow's tissues are supplied with plentiful acetate and relatively little glucose. Therefore the degradation of glucose is minimal and the milk-fat fatty acids are hardly labeled. Similarly, the ruminant does not waste laboriously synthesized glucose for combustion to carbon dioxide, whereas in the nonruminant, glucose is the chief fuel.

Ketosis. Ordinarily the ruminant is unaffected by its relative shortage of glucose because it is able to synthesize its requirements. In lactation, however, the high-producing cow is chronically on the verge of a serious shortage of this metabolite, and such a hypoglycemia appears to be the triggering factor in ketosis. Ketosis is an important metabolic disease of cattle remarkably similar to diabetes mellitus in man. It is true that the blood-glucose level in diabetes is elevated and in ketosis it is very low, but the effect on the cells is similar. In diabetes the deficiency of insulin causes a functional deficiency of glucose. In each case, the metabolism of acetyl coenzyme A units is impaired so that they accumulate and condense to yield keto acids. The excess keto acids cause the acidosis and acetonuria characteristic of both diabetes and bovine ketosis. Ketotic animals respond favorably to glucose injections which, of course, correct the hypoglycemia, but the ketosis returns within a few hours. The actual cause of the hypoglycemia is obscure but the disease is most common in heavily lactating animals, suggesting that the demands for lactose secretion outstrip glucose synthesis. Because butyrate and propionate are the common precursors of glucose, it is possible that the shortage is due to an impairment of the production, absorption, or metabolism of these fatty acids. It is quite certain that an insulin deficiency is not involved. *See* METABOLIC DISORDERS.

Ketotic animals will also respond favorably to ACTH injections, which suggests that pituitary or adrenocortical defects are involved. It is known, however, that ACTH potentiates the production of

glucose from protein breakdown, and some workers consider that the ACTH response is only that of a glucose injection.

In summary, fermentation in the rumen provides ruminants with a supply of short-chain fatty acids but no glucose. The tissues are adapted to using acetate as the main substrate for energy metabolism. Glucose required for lactose synthesis and other purposes is synthesized from short-chain fatty acids. The subsequent degradation of glucose to acetate or carbon dioxide is minimal. Ketosis is a metabolic disease of lactating cattle and is accompanied by hypoglycemia and acidosis. The peculiarities of carbohydrate metabolism in ruminants make them especially vulnerable to this disorder. *See* DIGESTIVE SYSTEM. [T.A.R.]

Metagalaxy

A synonym for universe, although more specifically the system of external galaxies or of extragalactic nebulae. The word metagalaxy is now seldom used in technical literature. The universe is the inclusive system of all the matter and radiation which exists. On the astronomical scale, this matter occurs in aggregate units as gas, dust, stars, clusters of stars, galaxies, and clusters of galaxies. *See* COSMOLOGY; GALAXY, EXTERNAL; GALAXY, THE; INTERSTELLAR MATTER; NEBULAE, GASEOUS; STAR. [A.R.S.]

Metagenesis

The phenomenon commonly referred to as alternation of generations, in which one generation of certain plants and animals reproduces asexually, followed by a sexually reproducing generation. Many organisms, such as the blue-green algae and certain protozoa, reproduce only by asexual means and do not exhibit metagenesis. This concept is more frequently associated with the plant kingdom, where life cycles occur in all major groups (divisions), than with the animal kingdom.

Plants. Two distinct generations, the gametophyte and sporophyte occur during the life cycles of plants which exhibit metagenesis. Sexual reproduction occurs during the gametophyte generation, as does the formation of the gametes, or sex cells. The cells of the gametophyte are haploid since they contain a single set of unpaired chromosomes. This stage in the life cycle begins with the germination of an asexual spore.

The formation of the zygote initiates the sporophyte generation. The zygote, or fertilized egg, is the result of the fusion of male and female gametes. Thus, this is a diploid structure having paired homologous chromosomes, with one set contributed by the male gamete and the other set by the female. The zygote develops into the sporophyte.

Among the various plant groups, there are great differences in both the degree of independence and prominence of the gametophyte and sporophyte generations.

Animals. Metagenesis is restricted to certain groups of animals, such as the coelenterates and some polychaete worms and urochordates, in which an asexual generation follows a sexual generation in a regular manner. The cladocerans, rotifers, and aphids exhibit an irregular alternation in which several asexual generations may intervene between a sexual generation. Environmental conditions, such as temperature, influence such alternations in cladocerans and may be a controlling factor in rotifers and aphids.

Both flukes and tapeworms may have complicated life histories in which successive generations occur. This can be considered to be metagenesis. However, if the various stages are thought of as larval, or developing, forms in which there is an increase in complexity until the adult stage is reached, the concept of metagenesis must be excluded. [C.B.C.]

Metal

An electropositive chemical element; that is, an element whose atoms form positive ions in solution. All chemical elements can be classified as either metals or nonmetals, although the distinction is not always sharp. About 75% of the elements are metals. Nonmetals are confined to the lighter elements of groups IVa through VIIa of the periodic table of the elements. *See* NONMETAL.

Some of the best known metals are iron, zinc, gold, silver, copper, platinum, mercury, lead, aluminum, and uranium. Less familiar metals include lithium, potassium, sodium, magnesium, barium, calcium, cadmium, manganese, chromium, cesium, cobalt, iridium, osmium, palladium, and molybdenum. The most abundant metals in the earth's crust are aluminum, iron, calcium, sodium, potassium, and magnesium. Studies of the sun's spectrum have revealed that these metals (in the gaseous state) are also the most abundant metals in the sun.

Physical properties. A metal is characterized physically by high electrical and thermal conductivity, luster, ductility, malleability, and (when polished) high reflectivity of light. Typical metals are crystalline solids, with crystal structures that are relatively simple and characterized by a dense packing and a high degree of symmetry. Most metals are of a grayish color, ranging from the bluegray of lead to the white of silver, although gold is yellow and copper is red. Densities of metals range from 0.53 g/cm³ for lithium (water = 1) to 22+ g/cm³ for osmium and iridium.

It is not true, however, that every metal possesses all the typical metallic properties mentioned in the preceding paragraph. For example, mercury is a liquid at ordinary temperatures. Tin exists in three solid allotropic forms at different temperatures—two (Sn_β, called white tin, and Sn_γ) definitely metallic, and the other (Sn_α, gray tin) having many of the properties of a nonmetal. An element which has the physical properties of a metal and the chemical properties of both a metal and a nonmetal is called a metalloid. Examples are arsenic, antimony, silicon, and tellurium. *See* METALLOID.

Chemical reactivity. The atoms of metals contain relatively small numbers of (negatively

charged) electrons in their outermost electron shells, while the atoms of nonmetals have a larger number. Sodium, for example, has a single electron in its outermost shell, while chlorine has seven.

Because of the arrangement of their electrons, metals show little tendency to form compounds with each other. They combine much more readily with nonmetals, for example, with chlorine to form chlorides (of which sodium chloride, just implied, is a common example), with oxygen to form oxides, and with sulfur to form sulfides.

Metals differ greatly in their readiness to form compounds, that is, in their chemical reactivity. The most active metals are lithium, potassium, rubidium, cesium, and radium. The arrangement of metals (and nonmetals) in a list of decreasing reactivity is called an electrochemical series.

Natural occurrence. The majority of metals are found in the form of mineral-bearing substances called ores. Gold, silver, copper, platinum, and bismuth, however, are often found in the free state (in native ores) because of their relatively low reactivity. Silver and gold were used for making ornaments at least 8000 years ago.

Alloys. One of the most important properties of metals is their ability to combine with other elements in stable solid mixtures known as alloys; most "metals" of commerce are actually alloys. Bronze, an alloy of copper and tin, was produced at least 5000 years ago. The field of metallurgy is concerned with the extraction of metals from their ores and with the preparation, structure, and physical properties of metals and alloys.

Electrical conductivity. The conduction of electricity in metals is carried out by motion of electrons only, in contrast to electrolytic conduction, which involves movement of both negative and positive ions. The best metallic conductor is silver, followed closely by the cheaper and more widely used copper. The term semimetal is sometimes used to refer to a metal that is a poor conductor of electricity.

The high electrical and thermal conductivities of simple metals, in particular those having only one electron in their outermost shell, can be satisfactorily explained by the free-electron theory of metals, in which the free electrons, that is, those giving rise to the conductivity, are regarded as moving in an electric potential which is approximated as constant throughout the metal. The free electrons are assumed to be identical to the valence electrons (the electrons in the outermost shells) of the free metal atoms. The conductivities (and other physical properties) of more complex metals, however, can be satisfactorily explained only by the more complicated band theory of solids, which represents one of the major triumphs of quantum theory.

The strange phenomenon of superconductivity, whereby the electrical resistivity of a substance essentially vanishes at temperatures near absolute zero, is exhibited by 23 metals, as well as by number of alloys and chemical compounds.

In this encyclopedia there are separate articles on each of the metals. For more detailed information on topics mentioned in this article and for other discussions pertaining to metals, see ACTINIDE ELEMENTS; ALKALI METALS; ALKALINE-EARTH METALS; ALLOY; ALLOY STRUCTURES; ATOMIC STRUCTURE AND SPECTRA; BAND THEORY OF SOLIDS; CHEMICAL BINDING; CONDUCTION (HEAT); CRYSTAL STRUCTURE; ELASTICITY; ELECTRICAL CONDUCTIVITY OF METALS; ELECTROCHEMICAL SERIES; ELEMENTS (CHEMICAL); FREE-ELECTRON THEORY OF METALS; METAL AND MINERAL PROCESSING; METAL COATINGS; METAL FORMING; METAL, MECHANICAL PROPERTIES OF; METALLOGRAPHY; METALLURGY; PERIODIC TABLE; PLASTIC DEFORMATION OF METAL; PLASTICITY; RADIOGRAPHY OF METALS; RARE-EARTH ELEMENTS; SOLID-STATE CHEMISTRY; SOLID-STATE PHYSICS; SUPERCONDUCTIVITY; TRANSITION ELEMENTS. [K.W.P.]

Metal, mechanical properties of

This article is a general discussion of the mechanical properties of metals that are commonly measured, such as the tensile properties, hardness, and fatigue strength. It is concerned with the way a metal behaves at a point subjected to various states of stress, and not with the general problem of the theories of elasticity and plasticity, which is the distribution of stress and strain throughout a body subjected to external loads. The latter is concerned mainly with the solutions of the differential equations prescribed by equilibrium. This article assumes that the state of stress at a point is known and discusses what is known experimentally about how the metal at this point reacts.

Deformation and fracture. Stress is defined as the intensity of force acting over an imaginary area in the immediate vicinity of the point at which the stress is to be described. It is the force per unit of cross-sectional area. A complete description of the state of stress at a point requires the magnitudes and directions of the force intensities on each surface of a solid surrounding the point. The number of such force intensity vectors and the analytical complexities are minimized when the point is placed at the center of a cube and when this cube is oriented so that the force intensity vectors are perpendicular to the cube surfaces rather than oblique. It is always possible to orient the cube so that all six of the boundary forces are perpendicular to the cube faces. The force intensity vectors on opposite faces describe what is called a normal stress when they are perpendicular to the surface. When the force intensity vector is parallel to the surface, the two forces on opposite faces describe a state of shear stress. Normal stresses are taken to be positive when they act to extend a line, that is, to produce tension. Shear stresses always occur in equal pairs of opposite sign at right angles. The

three normal stresses, which are called the principal stresses. The directions in which they act must be described and are known as the principal directions of stress. The boundary planes of the cube perpendicular to these directions are known as the principal planes of stress. In the general case, all other planes in the vicinity of a point are subjected to shear stresses. The shear stress is a maximum on the plane bisecting the dihedral angle between the principal planes on which act the greatest and least principal stress. The greatest normal stress in any direction is identical with the greatest principal stress. The magnitude and orientation of the maximum shear stress is of interest because inelastic deformation, whether viscous or plastic, occurs most rapidly on the plane and in the direction of the maximum shear stress, and this rate is strongly dependent upon the magnitude of the maximum shear stress. The magnitude and orientation of the maximum normal stress are of interest because fracture in general occurs on the plane perpendicular to the direction of the maximum normal stress and at a critical value of the maximum normal stress.

Tension and torsion. In simple tension, two of the three principal stresses are reduced to zero, so that there is only one principal stress. In simple tension, the maximum shear stress is numerically half the maximum normal stress. Because of the symmetry in simple tension, every plane at 45° to the tensile axis is subjected to the maximum shear stress. For other kinds of loading, the principal stresses and the maximum shear stress and the planes in which they act are different.

For instance, in simple torsion, the maximum principal stress is inclined 45° to the axis of the bar being twisted. The least principal stress is perpendicular to this, at 45° to the bar axis, but equal to and opposite in sign to the first principal stress, in other words, compressive. Both of these are in a plane perpendicular to the radial direction, the direction of the intermediate principal stress, which in this case has the magnitude zero. Every free external surface of a body is a principal plane on which the principal stress is zero. In torsion, the maximum shear stress occurs on all planes perpendicular to and parallel with the axis of the twisted bar; but because the principal [illegible] half of it, as in simple tension. This means that in torsion, one may expect more ductility (the capacity to *deform before* fracture) than in tension. Materials that are brittle (exhibiting little capacity for plastic deformation before rupture) in tension may be ductile in torsion, because in tension the critical normal stress for fracture may be reached before the critical shear stress for plastic deformation is reached; whereas in torsion, because the maximum shear stress is equal to the maximum normal stress instead of half of it as in tension, the critical shear stress for plastic deformation is reached before the critical maximum normal stress for fracture.

Another important situation arises at the root of a geometrical discontinuity in a loaded body. Such a discontinuity is referred to as a notch. At the root of a notch immediately beneath but very near the surface, the maximum normal stress may rise to magnitudes considerably greater than the maximum shear stress. This means that the maximum shear stress, instead of being equal to the maximum normal stress as in torsion, or half of it as in tension, may be a considerably smaller fraction. Also, the critical normal stress for fracture is more easily reached before the critical shear stress for deformation than in torsion and tension, and one may expect notches to have an embrittling effect. This embrittling effect of notches is of very great practical importance and accounts for the majority of service failures in engineering structures.

Brittle temperature. Many metals and alloys, of which unfortunately structural steel is one, exhibit what is known as a transition temperature, below which the metal is brittle and above which it is ductile (Fig. 1). This transition, or brittle, temperature is not a fixed temperature, such as is a melting point, but is the temperature below which the critical normal stress for fracture is reached before the critical shear stress for plastic deformation, and above which the reverse is true. This situation arises because the critical shear stress for plastic deformation is more sensitive to temperature than the critical normal stress for fracture; at high temperatures, deformation occurs at much lower stresses than fracture, whereas at very low temperatures, the fracture stress may be lower than the deformation stress. Some metals, all of them face-centered-cubic in crystal structure, have such a low deformation stress relative to the fracture stress, or such a slight dependence of the deformation stress on temperature, that a brittle temperature is never exhibited. Copper and aluminum are in this group.

This brittle temperature is different for different states of stress because the ratio of the maximum

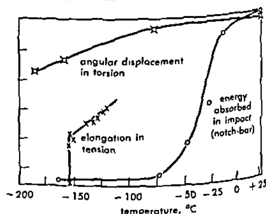


Fig. 1. Ductility in several tests as a function of the testing temperature. (*After J. H. Heindlhofer*)

shear stress to the maximum normal stress is different, as described above. The brittle temperature is lowest in torsion and highest in the presence of a severe notch. The brittle temperature for unnotched, common structural steel is quite near the absolute zero of temperature in torsion (certainly below 20°K and perhaps below the helium point, 4°K), just below the nitrogen boiling point in simple tension (about 60°K), but in the neighborhood of northern winter temperatures (250°K) in structures containing defects which can act as notches to concentrate the normal stress. All large structures contain such defects. When the structure is made continuous by welding, a single crack can break the entire structure. Large all-welded steel structures, therefore, present an important problem.

The brittle temperature for a particular state of stress depends on a number of factors, including the chemical composition and the microstructure of the material. The latter is affected by the physical dimensions of the structural element, in particular by the thickness of hot-rolled steel plates. Thick steel plates, probably because of their slower cooling rate from the rolling temperature, and perhaps because they have not been recrystallized in working as often as thinner plates, are characterized by higher brittle temperatures. Hence, as the size of engineering structures is increased, with an increase in the thickness of the plates used, the problem of brittle fracture becomes more serious, and it becomes necessary to use steels which, because of their chemical composition and method of manufacture, have lower brittle temperatures. Among the factors known to lower the brittle temperature of steel are (1) decreased carbon content, (2) increased nickel and manganese content, and (3) the use of a fine-grain steel-making practice (to minimize the grain size of the steel, an important factor). See ELASTICITY; PLASTIC DEFORMATION OF METAL; STRESS AND STRAIN.

Viscosity and strain hardening. In the preceding section, the resistance to plastic deformation was treated as though it varied only with temperature. It does, in fact, vary with the time-rate of straining and with the amount of strain. For modest variations of the time-rate of straining, the resistance to deformation changes little, but very large variations in the strain rate are obtainable, ranging from the very slow rates that characterize static loading to the very fast rates of straining that occur at the advancing edge of a moving crack in a brittle material. The strain at a fracture surface may reach several per cent, but there is an extremely steep strain gradient. The strain is reached in a very short time, since cracks have been observed to travel about 1 mi/sec. If it is assumed that the strain develops within 0.01 in. of the edge of the crack, this speed would lead to a strain rate of the order of magnitude of 10^6 sec^{-1} (in./in./sec). Such a high strain rate could elevate the resistance to deformation considerably. Steel with a resistance to deformation of about 40,000 psi at ordinary temperatures breaks at values of the normal stress of about 200,000 psi at very low temperatures, and at about the same value of stress by calculation at the root of a sharp notch.

Whereas fluids are usually Newtonian viscous bodies with a linear dependence of the shear stress on the time-rate of shear straining, metals exhibit a nonlinear dependence of shear stress on time-rate of shear strain. The time-rate of shear strain is proportional to an exponential function of the shear stress. One consequence of this behavior is that flow is not restricted entirely to the plane and direction of maximum shear stress, but occurs to some degree on other planes and directions. This modifies the simple rule for yielding in maximum shear stress to a rule requiring that an effective average shear stress be reached. This effective average shear stress may be called, in short, the effective stress, and by experiment turns out to be a function of the three principal stresses, known generally as the von Mises function, but also by other names describing the physical concepts it describes, such as the elastic shear strain energy, the octahedral shear stress, the root-mean-square shear stress, the mean crystallographically resolved shear stress for randomly oriented aggregates of cubic crystals, or simply the second invariant of the stress tensor. It may be written

$$\bar{\sigma} = \{\tfrac{1}{2}[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2]\}^{1/2}$$

where $\bar{\sigma}$ is the effective stress, and σ_1, σ_2, and σ_3 are the principal normal stresses.

Stress-strain curve. Normally, the viscous strain rate decays rapidly, more rapidly the lower the temperature, and becomes negligible in a few seconds. This decay of the strain rate results from the phenomenon of strain hardening. If a metal is strained at a fixed rate, the resistance to deformation increases with straining at a diminishing rate.

obtained by making a tension test and plotting the maximum normal stress against the maximum normal strain. These are identical with the ordinary tensile stress (the tensile force divided by the area of the specimen) and the ordinary tensile strain (the extension per unit of length). For small strains, the original area and the original gage length may be used without serious error in calculating the stress and strain, but where the strains are large, it is better to use the true average stress, obtained everywhere on the curve by dividing the tensile force by the actual area and the so-called true strain. The true strain is obtained for every point on the curve by integration of dl/l, between the limits of original length l_0 actual length l, thus taking into

changing gage length. The true strain is thus the natural logarithm of the ratio of the final length to the original length

$$\int_{l_0}^{l} \frac{dl}{l} = \ln \frac{l}{l_0}$$

The ratio of the final to the original length may be replaced by the ratio of the original area to the final area, because of the constancy of volume in inelastic deformation.

For metals which have not suffered any prior permanent deformation, the stress-strain curve in simple compression is usually nearly identical with the stress-strain curve in simple tension But this is not so for other states of stress. In torsion, for instance, the maximum normal stress reaches a value, at corresponding strains, of less than 60% of the value reached in simple tension. A more generally applicable stress-strain relationship is obtained by plotting the von Mises function of the three principal stresses (the effective stress) against an easily derivable function of the three principal strains that may be called the effective strain, and which turns out to be

$$\bar{\varepsilon} = [\tfrac{2}{3}(\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2)]^{1/2}$$

where $\bar{\varepsilon}$ is the effective strain and ε_1, ε_2, and ε_3 are the principal strains. Both the effective stress and the effective strain reduce to the simple tensile stress and tensile strain for simple tension testing, so that the ordinary tensile stress-strain curve without change in units may be used as the effective stress-strain curve (see Fig. 2).

Effective stress-strain curve. Unless some physicochemical change occurs in the metal during straining, such as recrystallization or precipitation of a second phase, the effective stress-strain curve usually is describable by the power function $\bar{\sigma} = k\bar{\varepsilon}^n$, where k is the effective stress required to produce unit effective strain, and n is an exponent that is always less than unity and commonly has a value between 0.1 and 0.4. The exponent n is called the strain-hardening index because differentiation of $\bar{\sigma} = k\bar{\varepsilon}^n$ to find the rate of strain hardening $d\bar{\sigma}/d\bar{\varepsilon}$ yields for this the expression $n(\bar{\sigma}/\bar{\varepsilon})$, so that the slope of the stress-strain curve (the strain-hardening rate) is proportional to n as well as to the ratio of the coordinates of the point on the curve. Higher values of n are associated with annealed pure metals and annealed single-phased alloys; lower values of n are associated with heat-treated alloys such as quenched and tempered steels.

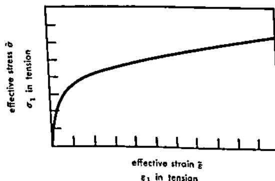


Fig. 2. Stress-strain curve.

Modern theory makes use of line imperfections (dislocations) to explain the low resistance to deformation of annealed metal crystals and polycrystalline aggregates, as well as strain hardening and viscous effects. Imperfection-free crystals deform only elastically and fracture at very high stress levels. This has been demonstrated with metal whiskers of high purity and of extremely small diameter. Larger crystals invariably contain line imperfections which move through the crystal lattice at low stress levels, but pile up at surfaces of contact with other crystals. Such imperfections are surrounded by stress fields which interact with each other, and strain hardening is ascribed to these interactions as one attempts to pile up larger numbers of imperfections at obstacles. Viscous effects are associated with the rate at which line imperfections can break loose from their obstacles, usually other line imperfections, with the help of thermal motion.

The effective stress-strain curve could in principle be determined in any test in which the forces applied and the strains produced could be measured. In practice, difficulties arise with every way of doing this. Many testing procedures are characterized by nonuniform distribution of the stresses and the strains with such steep gradients as to preclude the use of the test except as a qualitative indication of the resistance to deformation. Indentation hardness testing is an extreme example of this, and so to a lesser extent is torsion testing. Biaxial tension and compression pose obvious difficulties. Simple compression testing is difficult because, with short specimens, friction effects that appear as the specimen slides over the compression plates produce lateral stresses and raise the average pressure, whereas long specimens are subject to plastic buckling at relatively low strains Even the tension test, which is in general the most satisfactory, gives some difficulty because of the [illegible] with conse- [illegible] and strain. [illegible] sile test is [illegible] en is long enough to avoid end effects.

Tension test. To achieve uniformity of distribution of stress and strain in a tension test requires that the specimen be subjected to no bending moment. This is usually accomplished by providing flexible connections at each end through which the force is applied at a point on the centroid through the specimen in the gage length. The specimen is stretched at a controllable rate, and the force required to deform it is observed with an appropriate dynamometer. The strain is measured by observing

the extension between gage marks adequately remote from the ends, or by measuring the diameter and calculating the change in length by using the constancy of volume that characterizes plastic deformation. Diameter measurements are applicable even after necking down is begun. The elastic properties are seldom determined since these are structure-insensitive. Special refinements are necessary when the elastic stress-strain relationships are in question because the strains are very small and because elastic deformation is not characterized by constancy of volume.

Yield strength. The elastic limit is seldom determined. Metals are seldom if ever ideally elastic, and the value obtained for the elastic limit depends on the sensitivity of strain measurement. The proportional limit, describing the limit of applicability of Hooke's law of linear dependence of stress on strain, is similarly difficult to determine. Modern practice is to determine the stress required to produce a prescribed inelastic strain, which is called the yield strength. The amount of strain used to define the yield strength varies with the application, but is most commonly taken as 0.2% (a unit strain of 0.002 in. per in.). Because upon unloading the behavior is almost ideally elastic, it is possible to use the offset method of determining the yield strength from a plotted stress-strain curve; a line with a slope equal to Young's modulus is drawn, displaced from the stress-strain curve at low stress levels by the amount of strain used in the definition of the yield strength, and the stress at the intersection of this line with the stress-strain curve is taken as the yield strength (Fig. 3).

Tensile strength. The tensile strength is by definition the maximum tensile stress which a material is capable of developing, calculated by dividing the maximum load carried during a tension test by the original cross-sectional area of the specimen. It is therefore not the maximum value of the true tensile stress, which increases continuously to fracture and which is always higher than the nominal tensile stress because the area continuously diminishes. For ductile materials, the maximum load, upon which the tensile strength is based, is the load at which necking down begins. Beyond this point, the true tensile stress continues to increase, but the force on the specimen diminishes. This is because the rate of strain hardening has fallen to a value less than the rate at which the stress is increasing because of the diminution of area. If at any time during the extension of a specimen the force required to extend it diminishes, the strain will become localized. The condition for necking down is therefore that the force go through a maximum; that is, $\partial F/\partial l$ (the partial derivative of the force with respect to extension) goes through zero. When force and length are expressed as functions of stress and strain, it is easy to show that this condition is satisfied when the true rate of strain hardening falls to a value numerically equal

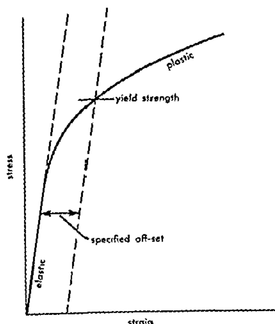


Fig. 3. Yield strength, most metals.

to the stress, that is, $\partial\sigma/\partial\varepsilon = \sigma$. For the stress-strain relationship used above, $\sigma = k\varepsilon^n$, this slope is reached when $\varepsilon = n$. In other words, necking down will occur when the tensile strain reaches a value equal to n (usually between 10 and 40% extension). The tensile strength for a ductile material has therefore nothing to do with fracture. It describes only the stress required to produce the tensile instability evidenced by necking down. For a relatively brittle material, one which fractures before the strain reaches the necking down value, the tensile strength reflects the fracture strength.

Yield point. A few alloys, notably structural steels, exhibit a yield point. Structural steel, because of the presence in solution of nitrogen and carbon, is characterized by an extraordinary elastic range. Up to stresses of the order of magnitude of the ordinary tensile strength, the inelastic strains are very minute. When the elastic limit is exceeded, the load instantly drops to a lower value at which straining proceeds by the movement of a plastic strain wave through the specimen. At the advancing edge of the plastic strain wave, the strain rapidly reaches the yield-point strain. During the period of advance of this strain wave, the force on the specimen is constant because the only plastically active region is at the advancing edge of the strain wave. The elastic limit is called the upper yield point, and the stress at which the plastic wave advances is called the lower yield point (see Fig. 4). During the progress of the plastic wave, its front is visible on the surface of the specimen because of the change in the thickness of the specimen at that point, and because the wave front leaves in its wake a somewhat roughened surface. This is spoken of as the Piobert effect, and the

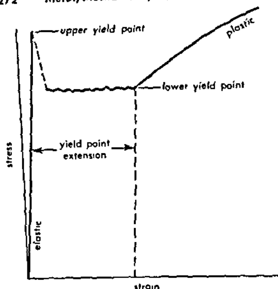


Fig. 4. Yield point, mild steel.

markings that are seen are referred to as Hartmann's lines or Lueder's lines. When the strain wave has completely traversed the specimen, the extension of the whole specimen is the same as that developed during the progress of the wave at the wave front. This extension is called the yield-point extension or yield-point elongation. This behavior is a characteristic of structural steels almost alone among engineering alloys, although similar marks are observed when the load drops during the tension test of other alloys because of structural changes that occur during the plastic extension.

Elongation. The tensile test provides a measure of ductility, by which is meant the capacity to deform by extension. The elongation to the point of necking down is called the uniform elongation because, until that point on the stress-strain curve, the elongation is uniformly distributed along the gage length (see Fig. 5 and compare with Fig. 2). The elongation to fracture or total elongation includes the extension accompanying local necking. Since the necking extension is a fixed amount, independent of gage length, it is obvious that the total elongation will depend upon the gage length, and will be greater for short gage lengths and less for long gage lengths. The American standard ratio of gage length to diameter for a cylindrical test specimen is 2:0.505 (the standard bar has a 2-in. gage length with a diameter of 0.505 in. corresponding to an area of 0.2 sq in.). The British standard ratio is 2:0.564, a diameter of 0.564 in. corresponding to an area of 0.25 sq in. The total elongation to fracture for the British specimen will therefore be more in the same gage length than for the American specimen. The European standard ratio is 5:1; elongations with the continental specimen will therefore be less. Absolute size does not affect the distribution of deformation. The shape of the cross section has no effect up to ratios of width to thickness of 5:1; rectangular cross sections elongate the same amount as circular cross sections of the same area. The reduction of area at the point of fracture, expressed usually as percentage reduction of area, is independent of gage length, and from it can be calculated the true strain at fracture; this makes the tensile reduction of area a more satisfactory measure of ductility than elongation unless one is interested in the amount that a metal will stretch before it begins to neck down. The reduction of area in the tensile test correlates well with measures of ductility obtained in other ways, as for example the strain to fracture on the outside of a bend, provided the strain gradient is not too great, as in notch-bar testing.

Notch-bar testing. The tensile test provides no indication of ductility in the presence of a sharp notch if, as described in the preceding section on deformation and fracture, the test temperature is below the transition temperature for the notched bar. Notched-bar tests are made to estimate the ductility that may be expected in the presence of defects in structures. Because the stress and strain distributions at the root of a notch are unknown, the results cannot be expressed in these terms. The common procedure is to measure the work required to break a standardized specimen, and to express the results in work units, as for example foot pounds. To express the results as foot-pounds of work per unit of cross-sectional area is misleading, because most of the strain and therefore most of the work done is concentrated at the root of the notch, and around the periphery where the crack meets the surface, so that the work varies more nearly with the linear dimensions of the specimen than with the cross-sectional area.

Notch-bar tests are usually made in either of two convenient arrangements, in both of which the specimen is broken by a freely swinging pendulum; the work done is obtained by comparing the position of the pendulum before it is released with the position to which it swings after striking and breaking the specimen (see Fig. 6). The specimen is struck by a tup placed at the center of percussion of the pendulum to minimize the reaction at the bearing about which the pendulum revolves. In the Izod test, the specimen is held in a vise, with

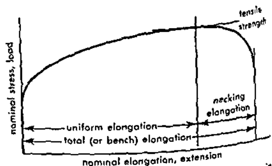


Fig. 5. Load-extension diagram (nominal stress strain curve) in simple tension.

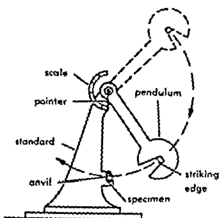


Fig. 6. Diagram of Charpy impact machine.

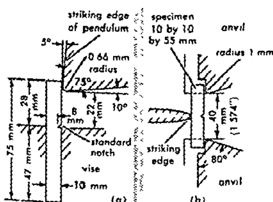


Fig. 7. Mountings of notch-bar test specimens for (a) Izod test, and (b) Charpy test.

the notch at the level of the top of the vise, and broken as a cantilever beam in bending with the notch on the tension side. In the Charpy test, the specimen is laid loosely on a support in the path of the pendulum and broken as a beam loaded at three points; the tup strikes the middle of the specimen, with the notch opposite the tup, that is, on the tension side. Both tests give substantially the same result with the same specimen unless the specimen is very ductile, a situation in which there is little interest (see Fig. 7).

Conditions of testing. The original Izod test used a sharper notch than the Charpy test. The Izod notch was a 60° V-notch, with a radius of curvature at the root of 0.01 in.; the notch was 0.080 in. deep; the specimen was 0.394 in. sq in cross section (1 cm sq). Modifications of the shape have been used, such as a circular cross section, but the notch has remained the same. The original Charpy test used a less sharp notch, parallel-sided; it had a radius of 0.039 in. and was 0.197 in. deep (1 mm radius, 5 mm deep, again in a 1 cm square bar). The Charpy notch is often made by drilling a 2-mm diameter hole and slotting to it with a saw; it is then called a keyhole notch. The results are the same as when the notch is made with a milling cutter, spoken of as a U-notch (see Fig. 8).

When testing at temperatures other than room temperature, the Charpy three-point loading arrangement is more convenient than the Izod arrangement, in which the specimen is clamped in a vise. It has been established that the temperature of the specimen will not change significantly at the root of the notch for several seconds, so that the specimen may be brought to temperature in a thermostatted bath and transferred to the testing machine for breaking. The use of the Izod V-notch in the Charpy test is increasing in popularity, and is known as the V-notch Charpy test.

The transition temperature obtained with the Charpy notch is somewhat lower than the transition temperature obtained with the more severe V-notch. The V-notch transition temperature correlates fairly directly with practical defects in structures, on the basis that brittle fractures in service are seldom observed when a specimen exhibits some ductility in the V-notch test. The number of foot-pounds that are needed to ensure freedom from brittle fracture in a structure depends on the material being tested, for the resistance to deformation as well as the ductility enters into the work required to break the specimens, along with the distribution of deformation which varies from one material to another. For example, ordinary plain carbon-steel ship-plate is not known to have been a point of origin of brittle fracture in ships when the steel exhibits more than 10 ft-lb energy absorption with a standard Izod notch. On the other hand, heat-treated alloy steel plates of lower carbon content may not be safe unless they exhibit perhaps as much as 50 ft-lb in the same test.

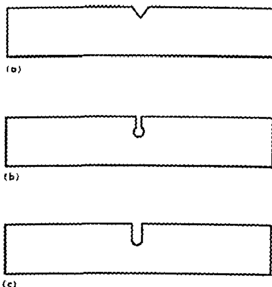


Fig. 8. Notch-bar test specimens with (a) V-notch, (b) keyhole notch, and (c) U-notch.

As in all tests involving brittle behavior, because brittle fracture is concerned with properties in a small region rather than the average throughout a large volume, considerable statistical variation is to be expected in notch-bar testing. Therefore, enough tests must be made to obtain a reliable result by statistical standards, especially when testing in the vicinity of the transition temperature.

Impact strength and stress level. The terms impact testing and impact strength are often applied to pendulum-machine notch-bar testing. This is not necessarily improper, for the rates of straining are about five orders of magnitude greater than in the ordinary tension test, but the rates of straining are still considerably lower than in true shock loading, and the results, including the transition temperature obtained when the specimen is broken in a pendulum machine, are not much different from when the specimen is broken in slow bending. Hence, it is the state of stress at the root of the notch that is important, not the speed of testing except to a secondary degree. In no test is there any control over the speed at which a crack propagates, once it has reached critical size.

The Griffith theory, which relates the critical size of a crack to the stress level, is applicable to metals, except that instead of the surface tension, which in glass is the only work to be supplied in making a new surface, one must use the total work required to produce the new surface. This total work, the specific energy of fracture, includes the work of plastic deformation near the surface and is of the order of magnitude of 100 in.-lb per sq in. or less for brittle alloys and 1000 in.-lb per sq in. or more for ductile alloys (above the transition temperature). The Griffith relationship is of the form

$$S^2 - S_0^2 = a\frac{E\omega}{\pi c}$$

where S is the nominal stress calculated by ordinary engineering formulas (the stress concentration at the crack is neglected), S_0 is a critical stress value established by the stress concentration factor and the deformation strength, E is Young's modulus, c is the depth of an edge crack or the half-width of an internal crack, ω is the specific energy of fracture mentioned above, and a is a constant of the order of magnitude of unity which depends upon the geometry of the situation; a is exactly unity for a crack in an infinite plate of finite thickness subjected to simple tension. Similar expressions have been worked out for the bending moment of notched bars in static bending, which may conveniently be used to evaluate ω. This provides the basis for a design procedure based upon the propagation of cracks instead of the conventional onset of plastic deformation.

Hardness testing. Tension testing provides a complete description of the relationship between stress and strain in plastic deformation, and in addition some useful information about ductility, and in conjunction with Griffith theory, some information about the resistance to crack propagation under conditions in which the metal is embrittled by the presence of a notch or crack. Often these methods of testing are more time-consuming and more expensive than is necessary, in particular when the only information that is needed is the comparison of the resistance to deformation of a particular sample or lot with a standard material. For such purposes, indentation hardness tests are used. They are relatively inexpensive and fast. They tell nothing about ductility and little about the relationship between stress and strain, for in making the indent the stress and strain are nonuniformly distributed.

In all hardness tests, a standardized load is applied to a standardized indentor, and the dimensions of the indent produced are measured. This applies to such methods as scratch hardness testing in which a loaded diamond is dragged across a surface to produce, by plastic deformation, a furrow whose width is measured, and the scleroscope hardness test in which an indent is produced by dropping a mass with a spherical tup onto a surface. The dimensions of the indent produced are proportional to the work done in producing it, and the ratio of the height of rebound to the height from which the tup was dropped serves as an indirect measure of the hardness.

Test methods. In the common Brinell hardness test, a hardened steel ball is forced into a surface by a force appropriate to the hardness of the material being tested. With steel and other hard materials, a 3000 kg force is applied to a 10-mm diameter ball; with softer alloys a 500 kg load is often employed. In every instance, the diameter of the indentation crater is kept between one-fourth and one-half the diameter of the ball (see Fig 9). The Brinell hardness number is defined as the force applied divided by the area of contact between the

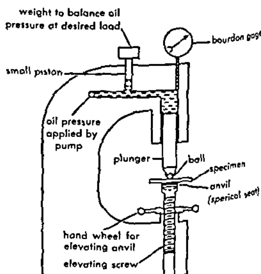


Fig. 9. Features of hydraulic-type Brinell machine.

ball and the test piece after the ball has been removed. The area of contact is calculated from the measured diameter of the crater after it has been assumed that the surface of contact is spherical and has the same radius of curvature as the unloaded indentor. The validity of this assumption depends, among other things, on the moduli of elasticity of the indentor and the test piece. It is usually incorrect but an incorrect assumption is of no consequence because the hardness is only relative in any event.

The Meyer hardness number is defined as the force applied divided by the area of the indent projected onto the plane of the original surface of the test piece. This number, just as useful as the Brinell hardness number, is easier to calculate but has not replaced the standard Brinell hardness number, the calculation of which is everywhere facilitated by easily available tables. A Meyer analysis of the Brinell test is carried out by plotting the logarithm of the applied force against the logarithm of the diameter of the indent. This plot is linear, and thereby demonstrates the validity of the relationship $F = ad^n$ where F is the force, d is the diameter of the indent, and a and n are the Meyer constants. Meyer's a is the intercept of the line at unit diameter (1 mm, normally), and Meyer's n is the slope of the logarithmic plot; it reflects the strain-hardening characteristics of the metal. It is possible to correlate any measure of strength, such as tensile strength, with any hardness number, provided the correlation is restricted to materials of similar strain-hardening characteristics, reflected in similar values of Meyer's n. The "ultimate Meyer's hardness number" is sometimes used in research papers to avoid the normal dependence of the hardness number on the force used. This dependence is due to the fact that a spherical indentor produces indents of different size that are not geometrically similar in shape and hence are characterized by different distributions of strain. The ultimate Meyer's hardness number is the force per unit of area of the projected indent, obtained by extrapolation in the Meyer analysis to an indent equal in diameter to the diameter of the indentor. This number is independent of the size of ball used and the loads applied to it.

The indentation-hardness number is independent of the force applied to the indentor when a conical or pyramidal indentor is used. The Martens' hardness test used a conical diamond indentor; it is no longer popular because of the fragility of the small-angle diamond cone that is used. A durable diamond indentor is used in the Vickers-penetration hardness testing method. It is a square-base diamond pyramid with a dihedral angle between opposite faces of 138°. This is the contact angle in the Brinell test when the indent is three-eighths of the ball diameter, in the middle of the range of indent sizes used in that test. This choice of angle yields hardness numbers in the Vickers penetration test about the same as those obtained in the Brinell test, with the advantage that the hardness number is independent of the load used down to quite small loads and quite small indents. Below loads of about 25 g, departures are observed. The Knoop hardness test is a modification using a diamond pyramid whose base is a rhombus, the diagonals of which are in the ratio of 7:1. The plastic deformation produced by this indentor is nearly plane strain, which can be accomplished in brittle materials without fracture in the adjoining regions under tensile stress, whereas cracks are produced by the nonplane strain produced by symmetrical indentors.

Rockwell hardness. The most popular hardness test is the Rockwell hardness test, which is carried out quite rapidly in a convenient machine by sacrificing any attempt to express the hardness as a resistance to deformation in units of force per area (Fig 10). Indentors of two shapes and various sizes are forced into the test-piece surface, first by a minor load, under which the position of the indentor is established as a reference point, and then by a major load which deepens the indent. Upon removing the major load and leaving the minor load applied, the amount by which the indent has been deepened is established. The hardness number is simply the amount of deepening on a linear scale, with the scale reversed so that soft materials having deeper indents are characterized by smaller numbers. The two shapes of indentors are (1) steel spheres of various diameters, for which the starting point on the linear scale for zero deepening is 130, and (2) a conical diamond with a spherical tip for which the starting point on the scale is 100. The hardness number is the number read from the dial gage indicating the depth of the indent on this reversed scale; it must be accompanied by a letter indicating the kind of indentor and load used. For example, for a load of 100 kg applied to a spherical steel indentor of ¹⁄₁₆-in. diameter, the letter B is used, so that a hardness number might be Rockwell B80; for the diamond indentor and a load of 150 kg, the letter C is used, so that a typical hardness number would be Rockwell C55. *See* HARDNESS SCALES.

Fatigue. When metals are subjected to cyclic straining, imperfections in the crystal grains accumulate, even though the inelastic strain in each

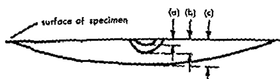


Fig. 10. Comparative impressions in steel (Rockwell C39) using Brinell, common Rockwell, and Rockwell superficial testers: (a) superficial Rockwell, N brale, 30-kg load, 0.0018 in., (b) common Rockwell, C brale, 150-kg load, 0.0052 in., (c) Brinell, 10-mm ball, 3000-kg load, 0.010 in. (From V. E. Lysaght)

cycle is very minute; and after enough cycles, fracture can begin at stresses which are considerably lower than those needed when the strain is monotonously increased to the fracture point. A failure which occurs after repeated application of a relatively low stress is called a fatigue failure. The number of cycles required to produce failure is a function of the stress level in each cycle. The results are expressed in the form of an *S-N* curve for which the stress or the logarithm of the stress is plotted against the logarithm of the number of cycles to cause failure. The fatigue strength is by definition the stress amplitude that will cause complete failure in a specified number of cycles, frequently 10^7 cycles Steels generally exhibit an endurance limit, beneath which it is not possible to produce fatigue failure, however often the stress cycle is applied (see Fig. 11).

Most fatigue testing is done with a mean stress of zero. A stress that varies sinusoidally with time is applied, sometimes in direct tension and compression but more commonly in bending by using a [illegible]

sequence, nor is the frequency at any easily achieved rotational speed. Interruptions in the testing are seldom of importance. The superposition of a static load upon the pulsating load has an important effect. The amplitude of pulsating *stress to* cause failure in a given number of cycles diminishes as the static or mean stress is increased. As in static testing under combined stresses, the maximum shear-stress or the von Mises function is the important stress function.

The cracks which form, sometimes at an early stage, progress a little each cycle until the crack is big enough to act as a Griffith crack and break the specimen entirely during one final cycle. The appearance of the fracture is different in the region of slow progress. When the stress amplitude varies,

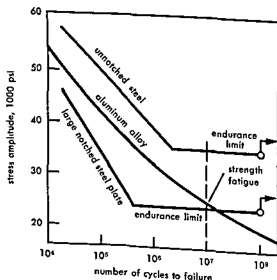


Fig 11. *S-N* curves, obtained during fatigue tests.

the rate of progress is different in different cycles, leading to striations upon the fracture surface that are sometimes spoken of as oyster-shell markings; when they are visible, they are a reliable indication that the failure is a fatigue failure. Statistical methods have been worked out to handle the problem of predicting the number of cycles to failure with a varying stress amplitude. Notches, even very rough surfaces, may significantly lower the fatigue strength, especially for the strong engineering alloys of high yield strength. The sensitivity to notches is less with more plastic alloys; in the limit, it is no more than would be expected from the elastic stress concentration caused by a notch. The notch-sensitivity factor q in fatigue is defined as

$$q = \frac{K_f - 1}{K_t - 1}$$

where K_t is the theoretical stress-concentration factor for ideal elastic behavior, and K_f is the experimental stress-concentration factor taken as the ratio of the fatigue strength of an unnotched specimen to the fatigue strength of a notched specimen Both K_t and K_f are always greater than *unity*. The value of q varies between 0 and 1 and is highest for strong alloys and lowest for annealed and soft alloys. Almost all fatigue failures start at notches, either notches countenanced by the designer or notches produced in manufacture such as machining defects and welding defects.

The fatigue strength is affected by prior mechanical history; it may be raised by very gradually raising the stress level as the specimen is cycled. The fatigue strength is of the order of magnitude of half the tensile strength, but is subject to considerable variation caused by such things as nonmetallic inclusions in the alloy. Corrosion, acting simultaneously with cyclic straining, can produce significant lowering of the fatigue strength. Residual stresses can modify fatigue behavior; tensile stresses act adversely and compressive stresses favorably. Fatigue life can be lengthened by the deliberate use of residual compressive stresses in the surface of a metal part, where fatigue cracks usually start. *See* METAL; METALLURGY. [M G]

Bibliography: American Society for Metals, *Metals Handbook;* ASTM Standards E8–54T, E6–54T, E23–47T, E10–54T, E18–55T, E48–47T; V. E Lysaght, *Indentation Hardness Testing*, 1949; W. M. Murray (ed.), M.I.T. Symposium, *Fatigue and Fracture of Metals*, 1952; R. E. Peterson, *Stress Concentration Design Factors*, 1953.

Metal and mineral processing

In the table are summarized the important steps followed in the extractive metallurgy of each of the known metals. For additional information, see the separate articles on each of the metals. *See also* ELECTROMETALLURGY; FLOTATION; HYDROMETALLURGY; IRON (EXTRACTION FROM ORE); LEACHING; METALLURGY; ORE DRESSING; PYROMETALLURGY, NONFERROUS; SEPARATION (CHEMICAL AND PHYSICAL); SEPARATION (MECHANICAL). [A.W.K]

Occurrence, extraction, and refining of metals (after M. F. Sittig)

Metals	Occurrence	Beneficiation operations	Extraction processes: Operation	Extraction processes: Intermediate product	Extraction processes: Product used in refining	Refining (reduction) operations
Aluminum, Al	Bauxite, Al_2O_3, 55–65% Al		Caustic (Baeyer process)	Sodium aluminate	Pure alumina	Fused salt electrolysis
Antimony, Sb	Stibnite, Sb_2S_3, 50–60% Sb	Liquation		Pure Sb_2S_3		Iron precipitation
Arsenic, As	Enargite, $3Cu_2S{\cdot}As_2S_5$		Roasting	As_2O_3		Charcoal at 1200–1300°F
Barium, Ba	Barite, $BaSO_4$	Flotation		Pure $BaSO_4$	$BaCl_2$	Fused salt electrolysis
Beryllium, Be	Beryl, $3BeO{\cdot}Al_2O_3{\cdot}6SiO_2$, 10–12% Be	Hand cobbing	Acid sinter plus leaching	BeO	BeF_2	Magnesium at 900–1300°C
Bismuth, Bi	Lead ores, 40–60% Bi	Flotation, gravity	Smelting	Lead bullion	Bismuth bullion by calcium treatment	Chlorine at 500°C to remove Pb and Zn
Boron, B	Borax, $Na_2B_4O_7{\cdot}10H_2O$	Evaporation		B_2O_3	Potassium fluoborate	Electrolysis
Cadmium, Cd	Sphalerite, ZnS, 1% Cd	Roasting, sintering	Leaching	Cd-Zn solution	Cd sponge by Zn dust addition	Distillation
Calcium, Ca	Limestone, $CaCO_3$		Calcination	CaO		Aluminum reduction in vacuum
Cerium, Ce	Monazite	Gravity	Sulfuric acid treatment	$Ce_2(SO_4)_3$	Tetravalent Ce salts	Electrolysis of fused chloride
Cesium, Cs	Lepidolite, 1% Cs	Hand cobbing	Sulfuric acid treatment	Cesium alum	CsCl	Calcium, aluminum, or magnesium reduction
Chromium, Cr	Chromite $FeO{\cdot}Cr_2O_3$, 35–50% Cr	Gravity	Electric furnace treatment	Ferrochrome		Electrolysis
Cobalt, Co	Cobalt or nickel ores, 0.3–10.0% Co	Flotation, leaching	Acid roasting or smelting	$CoSO_4$ solution	$Co(OH)_3$ by milk of lime	Electrolysis of aqueous solution
Columbium (niobium), Cb	Columbite, $Fe(Cb{\cdot}TaO_3)_2$	Gravity	Caustic fusion, then HCl then KF	$CbKOF_5$	Columbium oxide	Columbium carbide in vacuum
Copper, Cu	Copper ores, 0.4–6.0% Cu	Flotation, leaching	Smelting or electrolysis	Cu anodes	Copper cathodes	Electrolysis
Dysprosium, Dy	Monazite	Gravity	Sulfuric acid treatment	$Dy_2(SO_4)_3$	Pure salts by fractional crystallization and ion exchange	Electrolysis of fused chloride
Erbium, Er	Monazite	Gravity	Sulfuric acid treatment	$Er_2(SO_4)_3$	Pure salts by fractional crystallization and ion exchange	Electrolysis of fused chloride
Europium, Eu	Monazite	Gravity	Sulfuric acid treatment	$Eu_2(SO_4)_3$	Pure salts by fractional crystallization and ion exchange	Electrolysis of fused chloride
Gadolinium, Gd	Monazite	Gravity	Sulfuric acid treatment	$Gd_2(SO_4)_3$	Pure salts by fractional crystallization and ion exchange	Electrolysis of fused chloride
Gallium, Ga	Al and Zn ores	Caustic	CO_2 treatment in acid	Ga_2O_3 and $GaCl_3$		Electrolysis
Germanium, Ge	Sphalerite, ZnS, 0.01–0.015% Ge	Flotation	Roasting	Germanium oxide	Germanium chloride, then hydrolysis to pure oxide	Hydrogen reduction
Gold, Au	Elemental form, 0.001% Au	Gravity, flotation, leaching	Cyanide or smelting	AuCN or Au	Pure gold	Zinc reduction
Hafnium, Hf	Zircon, $ZrO_2{\cdot}SiO_2$	Gravity	Acid or alkali	Hf-Zr solutions	Hf solutions by ion exchange to $HfCl_4$, solvent extraction	Sodium reduction
Holmium, Ho	Monazite	Gravity	Sulfuric acid treatment	$Ho_2(SO_4)_3$	Pure salts by fractional crystallization and ion exchange	Electrolysis of fused chloride
Indium, In	Zinc ores	Flotation	Acid roasting	In sulfate	Water solutions	Displacement by zinc metal
Iridium, Ir	Cu-Ni sulfide ores	Flotation, roasting, extraction	Electrolytic refining	Anode slimes	$(NH_4)_2IrCl_6$	Hydrogen reduction
Iron, Fe	Iron ores, 30–60% Fe	Gravity, flotation	Blast furnace	Pig iron		Coke reduction in a blast furnace
Lanthanum, La	Monazite	Gravity	Sulfuric acid treatment	$La_2(SO_4)_3$	Pure salts by fractional crystallization and ion exchange	Electrolysis of fused chloride
Lead, Pb	Galena, PbS, 6–10% Pb	Gravity or flotation	Blast furnace	Lead bullion	Lead free of Cu, Sn, As, and Sb by treatment with sulfur	Desilvering with Zn, Sb removal
Lithium, Li	Spodumene, $Li_2O{\cdot}Al_2O_3{\cdot}SiO_2$	Flotation	Acid roasting	Li_2SO_4	Li_2CO_3, then LiCl	Electrolysis of fused chloride
Lutetium, Lu	Monazite	Gravity	Sulfuric acid treatment	$Lu_2(SO_4)_3$	Pure salts by fractional crystallization and ion exchange	Electrolysis of fused chloride
Magnesium, Mg	Seawater		Lime slurry	$Mg(OH)_2$	$MgCl_2$ with HCl	Electrolysis of fused chloride
Manganese, Mn	Manganese ores, 45–55% Mn	Flotation, gravity	Roasting in reducing atmosphere	Mn in solution	Pure Mn solutions by filtration	Electrolysis of aqueous solution

Occurrence, extraction, and refining of metals (after M. F. Sittig) (Cont.)

Metals	Occurrence	Beneficiation operations	Extraction processes: Operation	Extraction processes: Intermediate product	Extraction processes: Product used in refining	Refining (reduction) operations
Mercury, Hg	Cinnabar, HgS, 1–3% Hg	Sorting, screening	Roasting	Hg		Retorting ore at 700°C
Molybdenum, Mo	Molybdenite, MoS_2, 1–3% Mo	Flotation	Roasting	MoO_3	Mo by hydrogen reduction	Power metallurgy or arc melting
Neodymium, Nd	Monazite	Gravity	Sulfuric acid treatment	$Nd_2(SO_4)_3$	Pure salts by fractional *crystallization* and ion exchange	Electrolysis of fused chloride
Nickel, Ni	Nickel ores, 1–3% Ni	Flotation, leaching	Roasting, leaching	(NiCuFe)S or Ni	Pure NiS by molten Na_2S *extraction, or* Ni	Carbon reduction of NiO
Osmium, Os	Cu-Ni sulfide ores	Flotation, roasting, extraction	Electrolytic refining	Anode slimes	$OsO_2(NH_3)_4Cl_2$	Hydrogen reduction
Palladium, Pd	Cu-Ni sulfide ores	Flotation, roasting, extraction	*Electrolytic refining*	*Anode slimes*	*$Pd(NH_3)_2Cl_2$*	*Hydrogen reduction*
Platinum, Pt	Cu-Ni sulfide ores, 0.001% Pt	Flotation, roasting, extraction	Electrolytic refining	Anode slimes	$(NH_4)_2PtCl_6$	Hydrogen reduction
Potassium, K	Sylvite, KCl	Flotation, gravity				Sodium reduction
Praseodymium, Pr	Monazite	Gravity	Sulfuric acid treatment	$Pr_2(SO_4)_3$	Pure salts by fractional crys*tallization and* ion exchange	Electrolysis of fused chloride
Radium, Ra	Pitchblende	Gravity, flotation	Acid digestion	Ra sulfate	Radium bromide	
Rhenium, Re	Molybdenite, MoS_2	Flotation	Roasting	Re_2O_7	$KReO_4$	Hydrogen reduction
Rhodium, Rh	Cu-Ni sulfide ores	Flotation, roasting, extraction	*Electrolytic refining*	*Anode slimes*	*$(NH_4)_3Rh(NO_2)_6$*	*Hydrogen reduction*
Rubidium, Rb	Lepidolite	Hand cobbed	Acid roasting	Rubidium alum	RbCl	Ca, Al, or Mg reduction
Ruthenium, Ru	Cu-Ni sulfide ores	Flotation, roasting, extraction	Electrolytic refining	*Anode slimes*	*RuO_2*	*Hydrogen reduction*
Samarium, Sm	Monazite	Gravity	Sulfuric acid treatment	$Sm_2(SO_4)_3$	Pure salts by fractional crys*tallization* and ion exchange	Electrolysis of fused chloride
Scandium, Sc	Thortveitite, $(Sc,Y)_2Si_2O_7$		Roasting with carbon	Scandium carbide	Scandium oxide, then ScF_3	Electrolysis of fused chloride
Selenium, Se	Chalcopyrite, $CuFeS_2$	Flotation, roasting	Electrolytic refining	Anode muds	Sodium selenate	SO_2 precipitation from aqueous solution
Silicon, Si	Silica, SiO_2	*Flotation,* gravity	Smelting			Carbon reduction in electric furnace
Silver, Ag	Nonferrous ores, 0.07% Ag	Gravity or flotation	Blast furnace	Lead bullion	Ag-Zn mixture by Zn treatment	Zn distillation, electrolysis of Ag
Sodium, Na	Halite, NaCl	Flotation, leaching				*Electrolysis of fused* chloride
Strontium, Sr	Celestite, $SrSO_4$		Digestion with soda ash	$SrCO_3$	$SrCl_2$	Electrolysis of fused chloride
Tantalum, Ta	Tantalite, $Fe(Cb\text{-}TaO_3)_2$	Gravity	Caustic fusion, *then HCl, KF*	K_2TaF_7	Ta by sodium reduction	Vacuum sintering
Tellurium, Te	*Chalcopyrite,* $CuFeS_2$	Flotation, roasting	Electrolytic refining	Anode muds	Sodium telluride-tellurate	SO_2 from aqueous solution
Terbium, Tb	Monazite	Gravity	Sulfuric acid treatment	$Tb_2(SO_4)_3$	Pure salts by fractional crystallization and ion exchange	Electrolysis of fused chloride
Thallium, Tl	Sphalerite, ZnS	Roasting, sintering	Leaching of flue dust	Thallium chloride	TlS by H_2S treatment, then converted *to sulfate*	Electrolysis of aqueous solution
Thorium, Th	Monazite	*Gravity*	Sulfuric acid treatment	$Th(SO_4)_2$	Pure salts by fractional crystallization and ion exchange	Reduction of *ThF_4 by* calcium
Thulium, Tm	Monazite	Gravity	Sulfuric acid treatment	*$Tm_2(SO_4)_3$*	*Pure salts by* fractional crystallization and ion exchange	Electrolysis of fused chloride
Tin, Sn	Cassiterite, SnO_2, 1.5–5.0% Sn	Gravity, flotation	Reverberatory furnace	Impure tin		Electrolytic refining
Titanium, Ti	Rutile, TiO_2, ilmenite, *1–8% Ti*	Gravity, electrical	Chlorination	$TiCl_4$		Sodium *reduction*
Tungsten, W	Tungsten ores, 60–70% W	Gravity, flotation	Caustic fusion	Sodium tungstate	WO_3	Hydrogen reduction
Uranium, U	Uranium ores	Flotation, leaching	Acid or alkali *leaching*	U_3O_8	UF_4	Ca, Al, Mg, or Na re*duction*
Vanadium, V	Carnotite, $K_2O \cdot 2UO_3 \cdot V_2O$	Leaching	Roasting with salt	Sodium vanadate	V_2O_5	Calcium reduction
Ytterbium, Yb	Monazite	Gravity	Sulfuric acid treatment	$Yb_2(SO_4)_3$	Pure salts by fractional crys*tallization* and ion exchange	Electrolysis of fused chloride
Yttrium, Y	*Monazite*	Gravity	Sulfuric acid treatment	$Y_2(SO_4)_3$	Pure salts by fractional crystallization and *ion exchange*	Electrolysis of fused chloride
Zinc, Zn	Sphalerite, ZnS, *10–30% Zn*	Flotation, gravity	*Roasting*	ZnO		Retorting *with coke at* 1300°C
Zirconium, Zr	Zircon, $ZrO_2 \cdot SiO_2$	Gravity	Acid or alkali treatment	Hf-Zr solutions	HF removal by ion exchange, converted to $ZrCl_4$	Sodium reduction

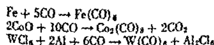
Metal carbonyl

A compound containing carbon monoxide combined with a metal. Iron and nickel carbonyls are liquids at room temperature; the others are crystalline solids which melt below 160°C. Metal carbonyls are decomposed on heating but can be volatilized at low temperatures and low pressures. They are insoluble in water but, in general, are soluble in organic solvents. The primary decomposition products are metals and carbon monoxide; however, metal oxides and metal carbides are formed under some conditions. Reaction with halogens, amines, bases, alkali metals, cyanides, phosphines, nitric oxide, acetylenes, olefins, aromatic compounds, and cyclopentadiene gives partial or complete replacement of carbon monoxide.

Metal carbonyls are formed only by the transition elements of groups VI, VII, and VIII of the periodic table. Those of iron, cobalt, and nickel are of technological importance.

$M(CO)_6$	Chromium, molybdenum, tungsten
$M_2(CO)_{10}$	Manganese, rhenium
$M(CO)_5$	Iron, ruthenium, osmium
$M_2(CO)_9$	Iron, ruthenium, osmium
$M_3(CO)_{12}$	Iron, ruthenium
$M_2(CO)_8$	Cobalt, rhodium, iridium
$[M(CO)_3]_x$	Rhodium, iridium
$M_4(CO)_{11}$	Rhodium
$M(CO)_4$	Nickel
$M_4(CO)_{12}$	Cobalt

Some metal carbonyls are prepared by direct combination of the metal with carbon monoxide; others by reduction of a metal compound in the presence of carbon monoxide. Some representative reactions for formation are


$$Fe + 5CO \rightarrow Fe(CO)_5$$
$$2CoO + 10CO \rightarrow Co_2(CO)_8 + 2CO_2$$
$$WCl_6 + 2Al + 6CO \rightarrow W(CO)_6 + Al_2Cl_6$$

Metal carbonyls are used in the preparation or purification of metals. An iron powder which has high magnetic permeability is prepared by decomposing iron carbonyl. Nickel is extracted from ores and simultaneously purified by the reaction of reduced ore with carbon monoxide to give nickel carbonyl. Its decomposition gives a powder composed of fine nickel spheres which is used in powder metallurgy. These and other metal carbonyls decompose on hot surfaces to give adherent, nonporous metal films. Some metal carbonyls also appear to possess significant antiknock properties in internal combustion engine fuels.

Reactions of carbon monoxide with olefins and acetylenes are catalyzed by metal carbonyls. Acrylic esters are prepared by the reaction of acetylene, alcohol, and carbon monoxide in the presence of nickel carbonyl (or with nickel carbonyl as the source of carbon monoxide). The hydroformylation reaction (reaction of an olefin with carbon monoxide and hydrogen to give a saturated aldehyde) is catalyzed by cobalt carbonyl. Iron carbonyl catalyzes the trimerization of acetylenes and the reaction of acetylenes with carbon monoxide to give quinones and cyclopentadienones. See HYDROFORMYLATION. [E.O.B.]

Metal coatings

Materials applied to the surfaces of metals to provide greater hardness, better wear resistance, better corrosion resistance, lubrication, or enhanced appearance. This article discusses hot metal coatings, metal spray and vapor coatings, plasma-spray coatings, flame-spray coatings, chemical coatings, and ceramic coatings. For discussions of other types of coatings, see ELECTROPLATING OF METALS; SURFACE COATING; SURFACE HARDENING OF STEEL.

HOT METALLIC COATINGS

Most metal coatings other than electrodeposits are applied hot to secure proper coverage and bonding. Of these the most common general methods are by hot dipping into a molten coating bath and by diffusion.

General hot dipping. This method consists of dipping a cleaned product through a molten bath of the coating metal, removing excess coating, and solidifying the covering film of metal. Preparation necessitates not only chemical or mechanical cleaning, but also protection of the surface with an active flux or nonoxidizing atmosphere until immersion in the coating bath. With few exceptions, such as in coating of steel with lead, some alloying occurs between the coating and base metal, assuring a tight bond. However, an excess of intermetallic compound at this interface promotes brittleness or flaking on forming. This is avoided by keeping the bath temperature as low as practical, by keeping the time of immersion short, and sometimes by adding alloying constituents to the bath. Excess adhering metal is drained off, shaken free, or even removed from some small objects by centrifuging. Rolls or fixed wiping mechanisms are commonly used to control the final coating on sheet, strip, and wire more effectively. Coating thickness may be as little as 0.00005 in. for lightly coated tinplate or up to several thousandths of 1 in. for heavy coatings of aluminum, zinc, or terneplate (a lead-tin alloy).

Galvanizing. This zinc-coating method protects by covering and also by galvanic or electrochemical action to prevent rusting over small exposed areas, such as the cut ends of wire. The importance of galvanizing is indicated by the 400,000–450,000 short tons of zinc used each year in the United States. Of this, nearly one-half is for the 2,500,000 tons of galvanized sheet steel produced; the rest is for tubes, wire, fittings, and miscellaneous products.

In conventional pot galvanizing, individual clean steel sheets or other products are passed through a flux mixture of zinc chloride and ammonium chloride before entering the molten zinc. Lead

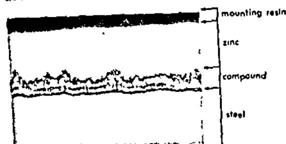


Fig. 1. Cross section of galvanized steel showing zinc-iron compound layer.

cadmium impurities in the zinc, and small amounts of antimony or tin which are sometimes added, help produce large, bright spangles. The method of cooling to secure selected points on the coating to start freezing also aids in spangle control. This is for appearance only; spangles are not an indication of quality.

By adding about 0.2% aluminum to the zinc bath, the ductility of the coating is greatly increased. Unfortunately, a chloride flux removes aluminum from the bath, and precautions must be taken to minimize this reaction or the aluminum content must be frequently renewed. Most sheet galvanizing is now done efficiently and continuously in single-strand, completely mechanized equipment, using aluminum in the zinc. One procedure is to pass the cleaned steel strip through a water solution of zinc and ammonium chlorides and to dry this salt on the strip before immersion in the zinc. Another, the Sendzimir type of line, dispenses with a chemical flux entirely. The steel strip is slightly oxidized to remove organic contaminants and is heated in a reducing atmosphere to give a clean active surface. Galvanized strip by either means may be used as such or cut into sheets. Much galvanized sheet is corrugated in a simple rolling operation to give greater rigidity for uses such as roofs and walls of buildings.

Sherardizing. This process differs from galvanizing in that it is entirely a dry diffusion operation. Tubes, castings, washers, bolts, nuts, and other steel parts are sealed in externally heated, rotating metal drums partly filled with zinc dust. In tumbling for 3–12 hours at 350–370°C, or well below the melting point of zinc, an intermetallic coating of 6–10% iron content is formed. The coating, usually 0.002–0.003 in. thick, is uniform, gray, and fissured with minute cracks which do not detract appreciably from corrosion resistance.

Aluminizing. This term designating the aluminum coating of base metals often refers to metal spray. However, the hot-dip aluminizing of steel sheet, wire, and miscellaneous formed articles for use at moderately high temperature is commercially important. Steel strip is commonly aluminized in equipment almost identical with that used in continuous galvanizing by the gaseous flux method. After a clean steel surface is assured by slight oxidation, then reduction of the oxide film in an atmosphere that is high in hydrogen, the strip passes through the molten aluminum bath and out through ports that prevent air contamination of the furnace atmosphere. By adding several per cent of silicon or several tenths of 1% of beryllium, the brittle alloy layer normally formed at the aluminum-steel interface is reduced sufficiently to permit easy fabrication.

Calorizing. In calorizing, or aliting as it is called in Europe, a diffused aluminum-iron coating is given to carbon- or alloy-steel products. Usually they are heated for 4–6 hours at 950–1050°C in a mixture of about 40% aluminum powder and 60% alumina containing a small amount of ammonium chloride. In a variation of this basic method, this pack treatment (or aluminum coating by metal spraying or by hot dipping) is followed with a diffusion heat treatment. In all cases, the aluminum content of the coating surface ranges from 25 to [illegible] accord- [illegible] eatment. [illegible] ions are [illegible] avoided. Applications include equipment for use in air up to 750°C or for shorter life at as high as 900°C, and for resistance to sulfur-containing atmospheres.

[illegible] is only one [illegible] vier coat- [illegible] metal may [illegible] be hot rolled over a properly prepared base metal (see CLADDING). Similarly, by rolling metal powder onto a metal base (compact rolling), followed by heating and usually by rerolling, a coherent clad surface of some metals is secured, for example, of copper, nickel, and stainless steel (see POWDER [illegible] is sometimes cast over a [illegible] opper is cast [illegible] le rolled and drawn to Copperweld wire, as examples of cast coatings.

In addition to using true metal vapor, usually in vacuum, as in evaporation and cathode sputtering techniques, many coatings of metals, alloys, and nonmetals can be made by chemical reduction of a volatilized compound of the coating material or by pyrolysis. Some coatings can also be applied by reactions in a molten salt bath. Coatings by these various means are usually designated after the coating metal, as chromizing, boronizing, siliconizing, and perhaps facetiously, tantalizing. [B W C.]

SPRAY AND VAPOR COATINGS

Solid objects may be given metallic coatings by spraying with molten droplets of metal or by deposition from a metal vapor. The former method is commonly used for relatively massive deposits and the latter for deposits of a few millionths of 1 in. in thickness.

Metal spraying. This method usually requires a spray booth from which fumes and gases are exhausted. The waste metal is often collected by a curtain of water at the rear of the booth, and by a tank of water covered with a grill on which the objects rest while being coated.

Low-temperature melting. For lead-tin solders, this operation is accomplished in a small electrically heated open pot or tank. The metal is drawn from the bottom of the tank (to avoid slag) by venturi action of an air jet which forcefully throws molten droplets of metal against the surface to be coated. One use is to provide solder-block terminals to the ends of metallized-paper capacitors.

Medium-temperature melting. For copper or steel, this operation is accomplished in a special torch called a Schooping gun. The metal in wire form is fed axially to an oxyhydrogen, oxyacetylene, or oxypropane flame, and the molten droplets are ejected with a continuous stream of compressed air (Fig. 2). A modification permits metals to be supplied in the form of powder instead of wire (Fig. 3). Worn shafting may be built up by the Schooping process and then remachined.

High-temperature melting. For tungsten and molybdenum, and refractory oxides, nitrides, and carbides, this operation is accomplished with a plasma-arc torch. As in the case of Schooping, materials may be supplied in either wire or powder form to the torch which makes use of a dc arc struck between a central tungsten cathode and a water-cooled copper anode. In some instances, the arc is struck between the tungsten cathode and the object being coated. In both cases, the molten splatter droplets are forcefully blown through the arc and ejected from the torch with a jet of inert gas. Most solid materials may be coated in this manner, including brass, some reinforced plastics, carbon, and graphite. It is claimed that the bonds are chemical and mechanical and that they are stronger than those obtained by electroplating, vapor depositing, and Schooping.

The plasma-arc torch has been used to coat the nose cones of missiles with tungsten. The coatings are very dense, usually laminar in structure, and may be finished to close tolerances.

Metal-vapor plating. This is the deposition of metal on a surface by means of condensation from a metal gas. The source of the metal vapor or gas

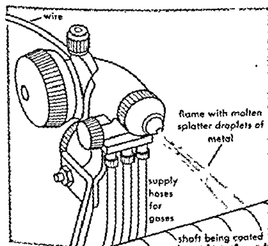


Fig. 2. Schooping gun.

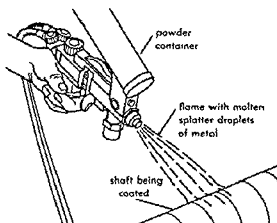


Fig. 3. Powder Schooping.

may be either evaporation or sublimation with heat in a vacuum chamber, or in a partially closed chamber at atmospheric pressure. The metal gas may also be provided by sputtering or pyrolysis.

The energy for gas formation in the various processes is provided by electrical-resistance heating, induction heating, electron bombardment, arcs, concentrated radiation as in a solar furnace, or as in the case of sputtering, energy transfer from accelerated ions. The atoms travel in straight lines for distances dependent upon the vacuum achieved. Not all the metal gas molecules stick to the surface of the recipient, but some are reflected or re-evaporated. The ratio of the number of metal atoms to the total number reaching the surface is the accommodation or sticking coefficient.

Vacuum heating. When a vacuum is used, the mean free path of the metal atoms should be large compared to the distance between the vapor source and the object to be coated. This usually means a vacuum of 10^{-4} mm Hg pressure or less. For batch coating, refractory-metal (tungsten) filaments are loaded with pieces of a lower-melting wire such as aluminum. In this way, aluminum is applied to toys, reflectors, and plastic automobile parts. These recipients are sometimes rotated for better distribution of the aluminum vapor. Front-surface spherical astronomical mirrors may be parabolized by rotation of a suitably shaped mask between the aluminum source and the mirror surface.

For semicontinuous metallizing in vacuum, long rolls of paper, plastic, or cloth may be coated from crucibles (usually carbon or ceramic material) containing molten metal. Wire or metal pellets are sometimes continuously supplied to the crucible to replenish metal evaporated.

Some uses of vacuum-deposited metal consist of electrodes for piezoelectric crystals, transistors, field-emission devices, and capacitors; printed-circuit patterns; and light reflectors in television tubes. Oxides, fluorides, and other stable metal compounds may be applied for use as nonreflective coatings for optical instruments and single and multilayer optical filters.

One novel use of metal-vapor plating is for high-speed electron beam recording, one set-up for which is shown in Fig. 4, with the vacuum bell jar raised. Untreated plastic film or lacquered paper from a supply roll passes in front of an open-ended cathode-ray tube *A*. With the bell jar closed and evacuated to 10^{-4} mm Hg pressure, the electron beam (terminating on a metal plate *B* behind the plastic) can write at linear speeds in excess of 500 ft/sec, leaving a latent image on the plastic or lacquer surface. This image may be developed by exposure first to an invisible monolayer of vapor-deposited silver (from *C*) followed by a semitransparent vapor deposit of zinc (from a source within the housing *D*).

Atmospheric heating. Cadmium and zinc have been experimentally evaporated and applied at atmospheric pressures to capacitor paper. The metal vapor is generated in and transported by a stream of inert or reducing carrier gas. Cadmium vapor is especially poisonous, and all fumes must be thoroughly removed.

Sputtering. Metal deposition by sputtering requires both a low-pressure controlled atmosphere (10^{-2}–10^{-6} mm Hg pressure range) and an electric field of several hundred to several thousand volts to produce a glow discharge. Resistance ballast in the circuit prevents arcing. The metal to be deposited is made the cathode (usually in wire or disk form), and the anode may be placed in a side arm near the vacuum pumps. The recipient is placed on a tray or rack usually within 1–2 in. of the cathode, and usually just outside the cathode dark space. Electrons from the cathode ionize the gas (usually argon) in the system. The ions are accelerated toward the cathode, where they strike and transfer their kinetic energy to atoms of metal, causing them to be ejected as a hot gas. This metal gas condenses on and coats any object in its path.

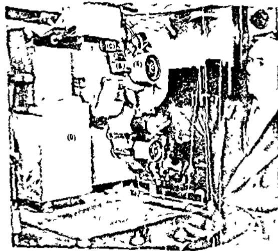


Fig. 4. Electron-beam recording. *(Bell Telephone Labs., Inc. and W. G. Matheson, ed., 1958 Fifth National Symposium on Vacuum Technology Transactions, Pergamon, 1959)*

The bulk of the cathode remains much below the melting point and hence much too cool to vaporize.

Sputtering requires an electric field to produce and accelerate ions to bombard a cathode, whereas vaporization depends upon thermal agitation of an entire surface to cause atoms to be ejected as a gas. Although sputtering is somewhat slower and more difficult to control than vaporization for production of uniform coatings over relatively large areas, it is especially useful for the deposition of highly refractory metals such as tungsten, tantalum, and molybdenum.

Pyrolysis. Pyrolytic metal-gas deposition is accomplished by bringing a gaseous metal compound into contact with a hot object to be coated. The heat causes decomposition of the compound and the metal is deposited on the object. A variation of this method involves introduction of hydrogen along with a volatile metal compound to free the metal by chemical reduction.

Gun barrels have been pyrolytically coated with molybdenum and with tungsten. Boron may be deposited on hot filaments for thermistor use. Carbon is often pyrolytically cracked from hydrocarbon vapor and deposited on ceramics for resistor use. Some additional refractory materials that may be pyrolytically deposited are as follows: tungsten from thermal decomposition of tungsten hexacarbonyl; titanium from thermal decomposition of the iodide or by hydrogen reduction of the tetrabromide; tantalum by hydrogen reduction of tantalum pentachloride; vanadium by thermal decomposition of the iodide; and molybdenum by thermal decomposition of the carbonyl or hydrogen reduction of the pentachloride. These and other metals also may be pyrolytically deposited as borides, carbides, and nitrides. [H.C.W.]

PLASMA-SPRAY COATING

The softening, or melting, of materials and propelling them in a molten, or plastic, state to a substrate, there to serve as a coating, is the technique commonly employed in plasma-spray coating. A variety of combustion processes has been devised, using such mixtures as oxyacetylene, oxypropane, and oxyhydrogen as the heat source. More recently, a completely new and different electric device, operating on a constricted-gas-stabilized-arc principle, has been introduced. This is popularly referred to as a plasma jet. The noble-gas plasma offers pure heat at temperatures up to 30,000°F. This temperature exceeds not only the melting points but also the boiling points of all the elements. Because the plasma is inert, no chemical change to the material transpires, as is possible with combustion processes. Although wire may be used, most types utilize powdered materials which are introduced into the plasma structure immediately downstream from the arc (Fig. 5).

Spraying, as performed by the plasma jet, is classified into six groups: electrical barriers (or opposite), radiation barriers, heat barriers, corrosion barriers, erosion barriers, and self-supporting

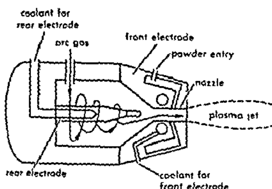


Fig. 5. Plasma-spray jet.

shapes. Within these classifications, it is possible to spray a very broad range of materials. However, because the art is very new, its full potential is yet to be established, and any list of capabilities is limited to the achievements to date and is subject to continuous change.

Considerable success can be expected from the oxides, including thoria and urania. Performing equally well are the borides, carbides, and to a limited degree, the nitrides. Stainless-steel alloys are particularly easy to manage. However, alloys sometimes contain low-melting-point elements, and care must be exercised lest these be burned out and the alloy be altered. Most of the elements and rare earths are manageable. Exceptions are aluminum and copper, whose melting points seem to be too low for the available temperatures.

Some method of surface preparation, usually a grit blast, is necessary. Because the bond is essentially mechanical, a roughened surface offers teeth to which the coating may cling. Substrate temperatures vary, but normal application does not excessively heat the material being coated. Although densities up to 96% of the pure metal are attainable, some porosity is present; this limits the use of coatings for corrosion resistance.

New concepts of fabrication are now practical, using the plasma spray. Crucibles and other more intricate shapes can be built up by continuous spray over a mandrel. Sandwiches of contrasting materials offer completely new material characteristics. Crystals have been formed with hardness equal to that of a diamond. The ability to design chemical or physical properties into a material for specific function is one of the most promising features of this industrial tool. See PLASMA PHYSICS. [B.LO.]

FUSED FLAME-SPRAY COATING

This is a method for applying hard-surfacing alloys in the form of finely divided metal powder. The powder deposit is fused to the base metal in a second step in the process, using the flame of an oxyacetylene torch. This fusing operation is the major difference between the flame-spray process and conventional metallizing procedures.

The flame-spray process is used whenever improved resistance to wear, corrosion, or galling is needed in component parts of industrial machinery and equipment. The resultant fused-metal alloy overlay provides even mild steel parts, for example, with better wear resistance than is normally found in high-grade tool and die steels.

Flame-spray coating is of greatest advantage when applied to cylindrical parts, although it is equally applicable to flat and irregularly contoured pieces. Among the more common parts typically hard surfaced using this method are shafts, sleeves, and rods used in a wide variety of rotating and reciprocating machinery. Cylindrical parts are best mounted in a lathe or turning fixture for both flame-spray and fusing operations.

Production of fused flame-spray hard-surface overlays involves three steps: preparing the base metal, spraying the powder onto the workpiece; and fusing the overlay to the base metal. The overlaid surface is usually finished by grinding following the fusing operation, although it can occasionally be used in the as-fused condition.

Process equipment consists basically of hopper, carburetor, pistol, hoses, and regulatory valves and gages. The pistol may be hand held, but is more frequently mounted in the lathe's toolpost to utilize the traverse mechanism. This permits constant-speed traverse during spraying, resulting in greater uniformity in application of the powder.

A regulated supply of oxygen, acetylene, powder, and compressed air is fed to the pistol through separate rubber hoses as shown in Fig. 6. [W.P.C.]

CHEMICAL COATINGS

The purposes of applying chemical coatings to metals are many and as varied as the types of coatings available. In general, such coatings are applied to provide a mechanical bond for paint or other organic finish, to retard corrosion, to retard wear, and to aid in cold forming of metals. See METAL FORMING; PAINT; PRIMER (SURFACE COATING).

These coatings are often very complex, and range from amorphous to crystalline in nature. Their wide use in industry has led to a simplification of terminology; they are now generally classified as zinc phosphates, iron phosphates, manganese phosphates, chromates, phosphate chromates, chromium-aluminum oxides, or oxalates.

No electric current is required to produce any of the coatings mentioned above. They are formed by a chemical reaction between the metal and chemically balanced solutions. The solutions may be applied to the metal by spraying or by immersion in a tank of the solution. The coatings produced are integral with the metal itself and hence have greater adhesion than any other known protective coating. The amount of coating produced is measured in milligrams per square foot of surface area. These so-called coating weights can vary considerably, depending upon the chemical solution, the type of metal or alloy, and the method of application. The end use determines to a great exte weight of coating desired.

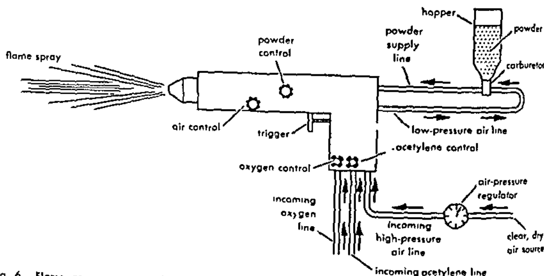


Fig. 6. Flame-spray process is of greatest advantage when applied to cylindrical parts, which are best mounted in a lathe or turning fixture. Flame-spray application of powder shown here is followed by a fusing operation.

Perhaps the greatest use of chemical conversion coatings is as a base for paint. For this purpose, zinc phosphates or iron phosphates can be produced on iron, steel, [illegible] alloys; c[illegible] mium-al[illegible] num and its alloys. Paint-base chemical coatings have several virtues in common: they are integral with the metal surface, provide a substantial mechanical bond with organic finishes, retard corrosion resulting from the small amount of moisture that penetrates the paint or other organic finish, and finally, prevent the spread of underfilm corrosion in the event the organic finish is damaged to expose bare metal. At relatively small cost, they serve to extend the life of applied finishes many times over.

The higher-quality paint-base chemical coatings are usually applied in 5- or 6-stage automatic spray machines, which incorporate cleaning, rinsing, and treating cycles at elevated solution temperatures. Some materials, however, are designed to clean and rinse metals in as few as one, two, or three stages. With fewer stages, good cleaning can be had, but very little phosphate coating is produced. Also tend[illegible] tions [illegible]

Gre[illegible] in the [illegible] the en[illegible] recogr[illegible] forme[illegible] has previously thought possible. In this application, the zinc phosphate is [illegible] used [illegible] the workpiece, even at the extreme temperatures and high unit pressures developed as a result of severe metal deformation. Consequently, less expensive raw materials can be utilized. Also, because of the amount of cold work performed, improved physical properties can be realized in the final part. Better lubricant qualities also result in less scrap, fewer operations, and less final machining.

Manganese phosphate coatings are used almost exclusively to retard wear on bearing areas and have found wide application in the treatment of gears, camshafts, valves, and cylinder liners. Their effectiveness lies in their ability to take and hold oil much as a blotter does while also serving as an antiflux, preventing metal-to-metal contact during break-in of moving parts. The relatively small build-up of coating, from 0.00015 to 0.0003 in. per surface, permits its use even in applications involving close tolerances.

The oxalate coatings as applied to lead, copper, or brass are of relatively minor importance, but their application in the cold drawing of stainless steel is spectacular. Oxalate coatings on stainless steel permit higher drawing speeds and heavier reductions, and at the same time produce smoother drawn surfaces.

Since mid-1958, new cold zinc phosphating and cold iron phosphating processes have been introduced and already have demonstrated dramatic savings in heat, water, and maintenance at the greatly reduced temperatures at which they operate. The cold processes represent one of the most important advances in the field in many years.

[R.W.E.]

CERAMIC COATINGS

These are inorganic materials applied to metals by a fusion process to improve the appearance, corrosion resistance, or other properties of the metal.

The coatings may be divided into two general types: glassy, which is usually designed for use at or near room temperature, and crystalline, which is more often used at high temperatures for its

refractory properties. Examples of the first type are the well-known enameled cooking ware, stoves, bathtubs, sinks, refrigerators, and signs, as well as the newer chalk boards and glass-lined tanks for breweries, dairies, and chemical-processing plants. Crystalline coatings are used to protect parts of jet engines and rockets exposed to high temperatures; a more unusual use is the coating of an earth satellite to control the emissivity of its surface and hence the temperature variations within as it moves in and out of the earth's shadow.

Glassy materials. Glassy coatings (known as enamels, vitreous enamels, or porcelain enamels) are generally silicate glasses with K_2O, Na_2O, B_2O_3, PbO, or Li_2O as fluxes (*see* GLASS AND GLASS PRODUCTS). Most enamels contain crystalline material, such as TiO_2, to make them opaque, but because the essential constituent is glass, they are classified as glasses. They are applied by two different processes, depending upon the type of metal base, the sheet-iron process and the cast-iron process.

Sheet metal, before being enameled, must be thoroughly cleaned; this is done by dipping it alternately into vats of alkali or acid and rinse water, the whole process being known as pickling. The clean metal is covered with the enamel, which consists of frit (powdered enamel), opacifier (a crystalline material which will not melt to glass in the firing operation), and suspending agents. Application is usually by spraying, although dipping, screening, or even hand painting (in the case of art jewelry) may be used. After the coating is dry, it is fired at about 800°C for a short time, a matter of minutes. *See* FRIT.

For enameling on aluminum, an even lower temperature must be used, and special enamels have been developed for this material.

For the best results, two coatings are applied. To obtain good adherence between metal and glass, the first or undercoat must contain one or more of the metals cobalt, nickel, or manganese. These metal ions impart a dark color to the glass which, for most purposes, must be covered by a second coat (the cover coat) containing an opacifier and the desired color, often white. Titania (TiO_2) is a common white opacifier. The clay used for suspending the enamel during application also adds to the opacity.

Cast iron for enameling is cleaned by blasting with sand or steel shot. The ground coat is sprayed on as in the sheet-iron process and the metal part (for example, a bathtub) is placed in a furnace until it is red hot. It is removed from the furnace by a long cantilevered fork lift, and while it is still hot, it is covered with frit by dusting from a box with a screen bottom. When the frit hits the hot piece it fuses, and a layer of molten glass is built up. All surfaces of the piece can be coated by tilting it at different angles. The quality of the enameling depends to a great extent upon the skill of the man doing the dusting. It is customary to apply the enamel in several steps, returning the casting to the furnace for reheating between times. In this way, flaws in the enamel coat can be repaired while the piece is still hot.

The glassy type of coating becomes molten at a temperature considerably below the melting point of the metal and therefore is not of great use at high temperatures. For these applications, the crystalline coatings were developed.

Crystalline coating materials. These are usually pure oxides (alumina, zircon, and stabilized zirconia), and are applied by the flame-spray method. A rod of the material is fed into an oxyacetylene flame which melts the material and blows it in droplets against the metal, where it flattens and solidifies. In this way, a coherent, solid, crystalline coating is built up on the surface. Before application of the coating, the metal surface is roughened by abrasion.

Because the oxide materials have higher melting points than the base metal, they can protect the metal against erosion and corrosion up to high temperatures, as well as acting as electrical and thermal insulators.

Another type of ceramic coating is formed on graphite by treating the carbon with vapor from silica sand (SiO_2) to form silicon carbide (SiC), a very hard, dense, refractory material.

The topic of ceramic coatings for metals is part of the general field of ceramic-metal combinations. It has become the practice in this field to use the term glass-metal when a glassy nonmetal is involved and the term ceramic-metal when a crystalline nonmetal is used. In all such work, one of the most important factors is the difference in thermal expansions of the two materials. In general, ceramic materials have lower expansions than metals, so that when they are applied as a coating to a metal they are placed in compression by the greater contraction (on cooling) of the metal. This is the desired state if the expansions of the materials do not match, because ceramics are relatively strong in compression and weak in tension. However, in ceramic-metal combinations such as large vacuum tubes, the stress system is more complicated, and tensile stresses are almost certain to be set up if the expansions of the metal and ceramic are not equal. This also applies to an intimate mixture of metal and ceramic, as in a cermet in which microstresses can occur as a result of unequal expansions. *See* CERMET. [M.C.M.]

Bibliography: I. E. Campbell, C. F. Powell, D. H. Nowicki and B. W. Gonser, Vapor-phase deposition of refractory materials, *J. Electrochem. Soc.*, 96(5):318-333, 1949; W. R. Cavanagh and R. C. Gibson, Phosphate coating of metal surfaces for industrial use, *Plating*, 42(6):742-748, 1955; V. M. Darsey, Effect of surface preparation on the durability of organic coatings, *Proc. Am. Electroplaters' Soc. Conf.*, 1946; V. M. Darsey and W. R. Cavanagh, Formation and application of phosphate coatings, *Trans. Electrochem. Soc.*, 91:351-365,

1947; L. Holland, *Vacuum Deposition of Thin Films*, 1956; H. S. Ingham and A. P. Shepard, *The Metco Metallizing Handbook*, 6th ed., 1954; J. F. Leland and J. W. Helms, *The Influence of Proper Lubrication in Design of Cold Extruded Components*, SAE Natl. Passenger Car Body and Materials Meeting, Paper 266, 1954; C. F. Powell, I. E. Campbell, and B. W. Gonser, *Vapor-Plating. The Formation of Metallic and Refractory Coatings by Vapor Deposition*, 1955; M. B. Roosa, Application of wear resistant phosphate coatings to ferrous surfaces, *Lubrication Eng.*, 6(3):117-121, 1950.

Metal forming

Manufacturing processes by which machine elements are fabricated from raw metal stock. Processes for metal forming may be classified arbitrarily into six major groups: casting; mechanical working such as forging, drawing, extrusion, rolling, stamping, and spinning; welding and allied processes; machining; powder and fiber metal forming, and electroforming. The selection of a process, or combination of processes, requires a knowledge of all possible methods of producing the part, if a serviceable part is to be produced at the lowest over-all cost.

Influence on service behavior. The service behavior of a machine element depends on the interrelation of the design, material, processing of the material (such as manufacturing process, heat treatment, and *surface treatment*), and operating and environmental conditions (such as temperature of operation, loading condition, and corrosive atmosphere). The *metal-forming* process is a factor in satisfactory service performance in that it affects the microstructure of the metal, it may introduce large residual stresses *into the* part, and it may affect the final design of the part which in turn influences its service behavior.

Process selection. The more important factors to be considered in choosing the optimum process (or combination of processes) include: type of material, metallurgical structure effect inherent in the process, size of the *part, shape* or complexity of the part, tolerances or finish required, quantity to be produced, cost, and production factors such as availability of equipment, *rate of* production, and time required to initiate production.

Material. The particular *metal or* alloy specified for a part is of major importance in the selection of the forming process. Some aluminum or copper alloys may be fabricated by practically any of the manufacturing processes; *other alloys* may be brittle under cold-working conditions but may be hot-worked. Highly refractory materials, such as tungsten and tungsten carbide, are not suitable for casting and must be fabricated *by* powder metallurgy methods. This process is also used for making porous metal products, or *parts requiring* combinations of two materials (not an alloy of the two, however). Alloys that are extremely hard, and therefore unsuitable for machining operations, can be precision cast and then ground if extremely close tolerances are required. Higher-melting point alloys such as steel can be fabricated by most of the major classes of processes but are not suitable for all of the individual processes within a major class. For example, steel can be sand cast but not die cast; it can be deep drawn but not readily impact extruded. Where a specific material must be used for the part, the choice of the optimum fabricating process may be definitely limited.

Metallurgical structure. Each manufacturing process has a different effect on the microstructure of the metal and consequently on its mechanical properties. Casting processes generally produce a relatively coarse-grained structure and a random orientation of nonmetallic inclusions. The result is isotropic properties, but lower ductility than that of wrought products. Castings may also be porous. Hot-working processes, such as forging, align the inclusions (fiber structure) and thus impart anisotropic properties with the strength and ductility generally being higher in a direction parallel to that of the inclusions. This orientation may be an advantage or disadvantage depending upon the direction of the applied loads. Cold-working processes (such as cold-rolling) also produce directional properties in the metal due to the tendency of the grains (or crystals) to align in certain directions. In addition the grains become distorted, and the metal becomes harder and stronger but less ductile. Cold-working operations (or any process causing nonuniform deformations) generally leave residual stresses in the part. For example, bending operations will, after release of the bending force, leave residual compressive stresses on the outer portion of the bends and tensile stresses on the inner surfaces. Residual stresses left in a part algebraically add to the stresses induced by service loading and in some instances are of *major* importance.

Size. Metal forming processes may be limited as to the size of the part that they can produce. Among the processes that are limited to relatively small parts are precision casting, die casting, powder metallurgy process, and screw machining. Large parts are best produced by sand casting, forging, or building up of component sections by welding or other joining processes.

Shape. The complexity of a part often dictates the process or combination of processes used in its manufacture. Generally, castings can be more complex than parts made by most other fabrication processes; however, some casting processes (sand and precision investment casting) are capable of producing parts of greater complexity than others. Parts produced by the powder metallurgy process have definite restrictions as to design because of the inability of metal powder to flow like a liquid. In some instances the complexity of a part *may* require the fabrication of the structure by welding or brazing several sections together. The change from forming a part as a single piece to fabricating it

from several sections usually requires design modifications if optimum serviceability is to be attained. The design should be left flexible until all feasible manufacturing processes have been considered.

Tolerances and finish. Parts requiring close tolerances or smooth finishes can be formed directly by precision investment casting, die casting, or such cold-working processes as swaging, drawing, or stamping. If formed by other processes, they can be finished by machining or grinding. Hot-working processes, such as forging, result in relatively rough, oxidized surfaces and relatively low dimensional accuracies. Welding operations generally result in some distortion or dimensional change. The final over-all cost often governs whether the desired tolerances should be attained in the original process or obtained in secondary operations.

Quantity. The number of parts to be produced is a major factor in determining the method of manufacture and whether the part should be produced within the plant or subcontracted. Some processes are suitable only for large-quantity production because of high tooling costs; for example, permanent-mold and die casting, certain forging processes, deep drawing, and screw machining. Processes such as sand casting, spinning, welding, and some machining methods are readily adaptable, but not necessarily restricted, to small-quantity production.

Cost. If the quantity to be produced is large, the over-all finished cost of the product is usually a prime consideration in the selection of a process. In many cases cost is the deciding factor in choosing the fabrication process and perhaps the material as well. In determining the over-all cost, the replacement cost of the part based on its estimated service life should be included with more immediate factors such as material and tooling costs, labor, and scrap loss.

Production factors. In some instances the time necessary to initiate production may be of significance in selecting the fabrication process. Those methods involving extensive tooling necessarily require a long lead time before production starts. Availability of equipment may be the deciding factor in choosing between two equally feasible processes, especially if it has been decided that the part is to be produced in a certain plant rather than subcontracting it to another organization. Another production factor which may be important is the required rate of production. Processes such as die casting, screw machining, powder metallurgy, and deep drawing have high production rates. Conversely, sand casting, spinning, hydraulic press forging, and fusion welding are relatively slow processes. *See* CASTING; DRAWING OF METAL; EXTRUSION; FORGING; MACHINING OPERATIONS; POWDER METALLURGY; ROLLING, METAL; SHEET METAL FORMING. [R.L.F.]

Bibliography: American Society of Mechanical Engineers, *ASME Handbook-Metals Engineering-Processes*, 1958; M. Begeman, *Manufacturing Processes*, 4th ed., 1957; H. Chase, *Handbook on Designing for Quantity Production*, 2d ed., 1950; E. P. De Garmo, *Materials and Process in Manufacturing*, 1957; J. L. Morris, *Modern Manufacturing Processes*, 1955; G. S. Schaller, *Engineering Manufacturing Methods*, 1953; J. F. Young, *Materials and Processes*, 2d ed., 1954.

Metal inspection, magnetic

Methods for the nondestructive testing of metal objects, depending upon the magnetic properties of the material. There are two widely used forms. One is called magnetic particle inspection. It readily detects cracks and flaws in iron and steel. It is so sensitive that it can uncover tiny surface defects or larger flaws deeply buried in the metal part.

The other test method, commonly called eddy-current testing, or more precisely, electromagnetic induction inspection, can measure such properties as composition, grain size, hardness, heat-treatment, internal stress, plating thickness, decarburization, or depth of hardening in all common metals. Modifications of the system can also measure wall thickness and disclose certain types of flaws.

Magnetic particle testing. If a part is magnetized and iron filings or iron oxide (magnetite) powder is sprinkled over it, a fine ridge of particles will adhere where there are surface flaws. With modern production equipment, the sensitivity of the method is increased by passing heavy current through the part or strongly magnetizing it in a coil energized by direct current. The iron dust is sometimes mixed with water or oil for easier application, and the addition of fluorescent powder aids in finding fine flaws when viewed under special light.

Many steel parts for automobiles, aircraft, and turbine engines must be highly heat-treated to obtain needed strength; this heat-treatment may

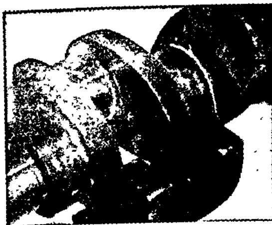


Fig. 1. Magnetic particle inspection shows hidden cracks in an automobile crankshaft before actual failure occurs. (Magnaflux Corporation)

cause cracks if not precisely controlled. Also, parts designed for unusually high working stresses may develop cracks in service. Magnetic particle inspection readily detects such cracks, even when they are still under the surface.

Eddy-current testing. In 1916, C. W. Burrows, a researcher at the National Bureau of Standards, discovered that the internal structure of a metal part influenced the energy it absorbed in a high-frequency field. He found that the method could detect instantly when any piece in a lot had a different composition, heat treatment, or grain size. During World War II, this method was used on a laboratory scale to inspect certain ordnance parts such as cartridge shells. A. B. Dumont developed the first practical commercial instrument in 1944, and J. W. Dice applied it in 1949 to automatic sorting and inspection for such common parts as nuts and bolts, gears, and piston rings. This instrument is widely used today for quality control and defect detection.

The eddy-current test is made by placing the part inside a coil powered by an electronic oscillator. The electromagnetic field sets up energy losses in the part, caused by magnetic hysteresis in the material and by the electrical eddy currents set up. These losses reduce the output of the oscillator and thus give a measure of these properties of the material. The principle is the same as for induction heating except the field is so weak it does not heat the part (see

ir

c

ti

sorting gates as the parts are passed through the coil.

Energy absorbed by magnetic hysteresis depends upon the molecular structure, permeability, and internal stress; any variation in these conditions will affect the oscillator output. Eddy-current losses are largely affected by electrical conductivity of the part, but the depth of penetration of the field depends upon the frequency. A low frequency will

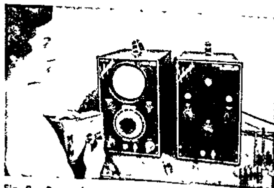


Fig. 2. Parts placed in the Cyclograph (eddy-current) test coil are rapidly sorted for quality. On the automatic sorter at right, three lights tell when measured quality is under, over, or within preset limits. (J. W. Dice Company)

penetrate deeply; a high frequency will stay on the surface. A low frequency (2–10 kc), by fully penetrating the part, gives an indication of the over-all structure, composition, or internal stress. A medium frequency (10–50 kc) gives similar information for the portion between center and surface. Higher frequencies (60–200 kc) measure conditions in surface layers such as case depth, decarburization, plating thickness, or surface stresses.

Certain types of cracks or flaws also can be detected. In a part with high internal stresses, a crack increases stresses around it, and this condition is easily detected. Easiest to find are radial cracks in a ring-shaped part or longitudinal cracks in a tubular part. Internal voids are difficult to find unless the volume is 10% or more of the part.

Although test indications are highly sensitive, they are, in general, relative and are most useful in comparing similar parts. In applying the method test coils are selected for the desired frequency and shape of part, and sensitivity controls are adjusted for the widest difference in indication between good and bad parts. Then, unknown parts are inserted in the coil and classified according to the indications given by the instrument. When testing large quantities, the operation can be made completely automatic through relays and sorting mechanisms with rates as high as five pieces per second for light, small parts. Automatic operation is particularly valuable in sorting mixed lots or incorrectly processed parts.

On a somewhat more complex set-up, bar stock and tubing are checked continuously in a steel mill. Two coils are used in a bridge circuit; automatic controls inspect the quality of the material. Similar equipment also locates any defective areas and marks them as each tube or bar is passed through the inspection equipment at high speed. See METAL INSPECTION, ULTRASONIC; METALLOGRAPHY.
[R.W.CA.]

Metal inspection, ultrasonic

The application of ultrasonic vibrations to materials with elastic properties and observation of the resulting action of the vibrations in the material. Ultrasonic frequencies are generally considered to be in the range above 18 kilocycles per second (kc), frequencies above the response range of the human ear. The frequencies usually applied to the testing of metals are in the range between 500 kc and 25 megacycles (Mc), although frequencies as low as 25 kc and as high as 200 Mc have been applied to special applications. These are high-frequency mechanical vibrations, similar to audible sound waves, and are various modes of particle vibration. They travel best in most liquids and metals and are attenuated rapidly in air or gases.

The important characteristics of ultrasonic testing are that the vibrations travel long distances through metals, up to 100 ft or more, in a well defined beam form and are reflected at disconti-

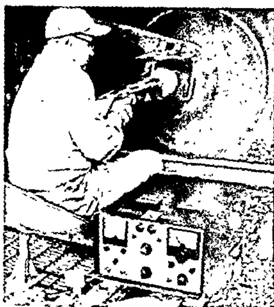


Fig. 1. A single frequency, 2¼-Mc ultrasonic test instrument for rapid repetitive-type testing, extensively used for locating fatigue cracks in the journal area of freight-car axles. A red light or buzzer automatically alerts the operator when a crack is located. A special transducer assembly of 12 quartz crystals mounted in a circle makes an instantaneous test when the probe is placed against the end of the axle. More than 2000 axles may be tested in one day with this ultrasonic equipment. (From Curtiss-Wright Corporation, Princeton Division, Princeton, N.J.)

nuities or boundaries of materials having different elastic and physical properties. Also, the mode of vibration may change when passing between materials having these different properties or when reflected from boundaries at an angle.

Ultrasonic inspection and testing will furnish three types of information about metals: location and extent of surface and internal discontinuities or nonhomogeneous areas in the metal, differences in structure and physical properties, and thickness (by introduction of vibrations from only one side).

Ultrasonic vibration generation. Of the modes of ultrasonic vibrations that may be generated in metals, there are four that are used for nondestructive testing and inspection:

1. Longitudinal (compressional) waves with longitudinal particle motion in the direction of propagation (like compressional audible sound waves).
2. Transverse (shear) waves with transverse particle motion in a direction perpendicular to the direction of propagation (like torsional waves).
3. Surface (Rayleigh) waves, propagated on the surface with particle motion a composite of compressional and shear (like waves traveling over the surface of a liquid).
4. Plate (Lamb) waves generated in thin metal, such as in the form of sheet having a dimension near or less than 1 wavelength; usually a variety of simultaneous modes having different velocities.

The velocity of ultrasonic vibrations (longitudinal, transverse, and surface modes) is dependent upon the density and dynamic elastic properties of the material. Wavelength, velocity, and frequency are related by the equation $\lambda = V/f$. The velocities of these three wave forms can be formulated as functions of the elastic moduli. Interrelated are Young's modulus of elasticity, shear modulus, bulk modulus of elasticity, Poisson's ratio, specific gravity, and velocity, which provide a basis for determination of the dynamic properties of metals. Elastic moduli and velocities can be interrelated as functions of Poisson's ratio only. See ELASTICITY.

Inspection methods. Methods of applying ultrasonic inspection of metals may be divided into three general categories.

In the pulse type, short pulses of ultrasonic vibrations are projected in a beam form into a metal; the energy reflected from discontinuities and the opposite side provides a basis for measuring the elapsed time of the round trip of the pulse and a calibration for distance in or thickness of the test piece. The pulse type is also used for through transmission testing.

In the continuous-wave method, the energy transmitted through specific areas of metal is measured.

In the resonance method, a continuous beam of ultrasonic vibrations is projected through a piece of metal. Resonance will occur when the thickness of the metal is equal to one-half the transmitted wavelength. A metal thickness can then be determined by finding the frequency at which resonance occurs.

A typical pulse-type ultrasonic testing instrument contains the following electronic circuits: synchronizer, pulser, sweep, marker, amplifier, power supply, and cathode-ray tube on which the test information is presented.



Fig. 2. An automatic immersed ultrasonic test being made on forged jet-engine rotors. Forged jet wheel is being lowered on to the turntable jig for testing. (From Curtiss-Wright Corp., Princeton Division, Princeton, N.J.)

Nozzle condensation and recombination. In actual engines, a certain fraction of the useful energy which is lost by vaporization and dissociation is recovered by condensation and recombination in nozzles. If thermodynamic equilibrium prevails, and if the nozzle area ratio is infinite, all of this energy loss is recovered. In actual cases, however, because of the finite area ratio and finite reaction and condensation rates, only a part of the energy is recovered for propulsion.

Loss due to condensed phase. A second type of propulsive efficiency loss occurs as a result of the appearance of a condensed phase. Because the acceleration of the gases in the nozzle results from pressure-volume work, a reduction in gas content reduces the thermodynamic efficiency. The combustion gas behaves as if the molecular weight were increased. A second consideration is involved with the momentum exchange between condensed phase and gas. If the particles are large and lag appreciably behind the gas in velocity, impulse is lost. This effect can be appreciable when the percentage of condensed phase is high.

The net result of the thermodynamic limitations is to reduce partly the improvement in performance expected on the basis of heat of combustion alone, but the properties of metal-base fuels still remain attractive. As an additive, metals are most effective when mixed with the lower-performing propellants because the lower temperatures minimize the vaporization and dissociation losses. For this reason [illegible] of such additives has been the closing of the gap in performance between the solid and liquid propellants. [D.A.]

Bibliography: J. V. Charyk and M. Summerfield (eds.), *High Speed Aerodynamics and Jet Propulsion*, vol. 2, 1956; M. Gilbert, L. Davis and D. Altman, Velocity lag of particles in linearly accelerated combustion gases, *Jet Propulsion*, 25(1):26–30, 1955; W. R. Maxwell, W. Dickinson and C. F. Caldin, Adiabatic expansion of a gas stream containing solid particles, *Aircraft Engineering*, 18:350–351, 1946.

Metallic-disk rectifier

A rectifier that consists of one or more disks of metal in contact with coatings or layers of a semiconductor material. Alternating current is changed into pulsating direct current by the rectifying action that occurs at the junction interface between a metal disk and its mating semiconductor layer. The most common examples are the selenium and copper-oxide rectifiers. In a selenium rectifier a thin layer of selenium is deposited on one side of an aluminum plate, and a conductive metal coating is sprayed or otherwise deposited on the selenium; electrons flow much more readily in the direction from the metallic coating to the selenium. In a copper-oxide rectifier, rectification occurs at the junction between a copper disk and a coating of cuprous oxide. See SEMICONDUCTOR RECTIFIER. [J.M.R.]

Metalloacid elements

The chemical elements with the following atomic numbers and names: 23, vanadium, V; 41, niobium, Nb; 73, tantalum, Ta; 24, chromium, Cr; 42, molybdenum, Mo; 74, tungsten, W; 25, manganese, Mn; 43, technetium, Tc; 75, rhenium, Re. These elements are the subgroup members of groups V, VI, and VII, respectively, of the periodic table. (See PERIODIC TABLE.) In the elemental state, all are metals of relatively high density, high melting point, and low volatility. The classification metalloacid elements refers to the fact that their oxides react with water to give somewhat acidic solutions, in contrast to the more typical behavior of the oxides of other metals which yield basic solutions.

The acid character of the oxides may be correlated with high positive charges attained by the metal ions in these compounds. (Maximum ionic charges attainable are equal to the respective group numbers, 5, 6, and 7.) In aqueous solutions, these ions repel positively charged hydrogen ions and attract negatively charged oxide or hydroxide ions from water molecules in their vicinity. The resulting solutions contain free hydrogen ions and therefore are acidic.

With the exception of technetium, which (because it has no suitably long-lived isotope) probably does not occur naturally in the terrestrial environment, the elements are quite important commercially. Because of its high melting point (3370°C) and low volatility, tungsten has long been used to form the filaments of high-temperature incandescent lamps. Vanadium, chromium, manganese, and niobium are essential components of various steels. The resistance of molybdenum and tantalum to chemical corrosion recommends their application in a variety of industrial processes.

Although technetium is probably not found as a naturally occurring terrestrial element, its spectrum has been observed in certain types of stars. The element is available in small amounts as a product of nuclear fission. It has potential usefulness in the treatment of metal surfaces to inhibit corrosion. See TRANSITION ELEMENTS. [B.B.CU.]

Metallography

The study of the structure of metals and alloys, especially by optical microscopy. In older usage, the term metallography designated the science relating the properties of metals to their structure, but this is now commonly called physical metallurgy. See METALLURGY.

The [illegible] metals is carried out [illegible] red to [illegible] agnifi- [illegible] higher

Microstructures of uranium-carbon alloys. (a) 4.46% C; (b) 4.86% C; (c) 5.75% C; (d) 7.98% C, (e) 9.08% C; all ×50. Samples were prepared by arc melting uranium and carbon in an argon atmosphere on a copper-cooled hearth. The ground and polished surfaces were chemically stained to produce a thin oxide film, the thickness and resulting color of which is dependent upon the crystallographic orientation of the substrata. The color indicates the grain size and amount and location of the different phases. *(USAEC-Union Carbide Corp. Oak Ridge Natl. Lab.*



(a)



(b)

(c)

(e) (d)





fining, and copper matte production all use reverberatory furnaces but have few other similarities. The physical metallurgist who is familiar with aluminum and its alloys may know little about steel or magnesium. The field of metallurgy is far from devoid of general principles but their applications require intimate knowledge of the characteristics of the individual metals. *See* ALLOY; ELECTROMETALLURGY; ELECTROPLATING OF METALS; HYDROMETALLURGY; IRON (EXTRACTION FROM ORE); METAL COATINGS; METAL FORMING; METAL INSPECTION, MAGNETIC; METAL INSPECTION, ULTRASONIC; METALLOGRAPHY; ORE DRESSING; PYROMETALLURGY, NONFERROUS; STEEL MANUFACTURE; see also articles on individual metals. [E.R.JE.]

Metamerism

The serial repetition of parts along the longitudinal body axis in bilaterally symmetrical animals. It is also called segmentation. The successive subdivisions are called segments, somites, or metameres. Common examples are the muscles and spinal nerves in the human body and in the body and tail of many mammals, snakes and lizards, salamanders, and fishes. If repeated parts are essentially alike, as along the body of a snake or earthworm, the segmentation is homonomous. When unlike, as in somites of the crayfish or insects, it is termed heteronomous. In the heads of arthropods, several primitive embryonic somites become fused in the adults. [T.I.S.]

Metamerism, embryonic

The repetitive formation of similar structures along the length of the body axis. It never involves reproductive organs, and thus differs from strobilization in tapeworms and certain jellyfish. True metamerism occurs in annelid worms, arthropods, vertebrates and other chordates. It originates either from a bilateral series of coelomic pouches which form the segmental muscles, kidneys, and body cavities of lower forms, or from mesoblastic somites which form the skeletal and muscular segments of vertebrates. Repetitive features of the nervous system are acquired secondarily through the influences of mesodermal metameres upon adjoining ectodermal tissues. *See* COELOM; EMBRYOLOGY; NEURULATION. [H.L.H.]

Metamorphic rocks

One of the three major groups of rocks that make up the crust of the earth. The other two groups are igneous rocks and sedimentary rocks. Metamorphic rocks are pre-existing rock masses in which new minerals, or textures, or structures are formed at higher temperatures and greater pressures than those normally present at the earth's surface. *See* IGNEOUS ROCKS; SEDIMENTARY ROCKS.

Two groups of metamorphic rocks may be distinguished: [illegible]lastic rocks, formed by the

formed under the influence of metamorphic pressures and temperatures.

Cataclastic rocks are mechanically sheared and crushed. They represent products of dynamometamorphism, or kinetic metamorphism (*see* METAMORPHISM). Chemical and mineralogical changes generally are negligible. The rocks are characterized by their minute mineral grain size. Each mineral grain is broken up along the edges and is surrounded by a corona of debris or strewn fragments (mortar structure, Fig. 1*a*). During the early stages of this alteration process the metamorphosed product is known as flaser rock (Fig. 1*b*). Eventually the original mineral grains are entirely gone, as in the mylonites. When seen through the microscope the comminuted particles consist of a mixture of finely powdered quartz, feldspar, and other minerals with an incipient recrystallization of sericite or chlorite. Pseudotachylite is an extreme end product of this crushing process. *See* FLASER ROCK; MYLONITE.

Structural relations. Metamorphic rocks, properly so called, are recrystallized rocks. The laws of recrystallization are not the same as those of simple crystallization from a liquid, because the crystals can develop freely in a liquid, but during recrystallization the new crystals are encumbered in their growth by the old minerals. Consequently the structures which develop in metamorphic rocks are distinctive and of great importance, because in many ways they reflect the physicochemical environment of recrystallization and thereby the genesis and history of the metamorphic rock.

Crystalloblastic structure. A crystalloblast is a crystal that has grown during the metamorphism of a rock. The majority of the minerals in metamorphic rocks are irregular in outline (xenoblasts), but some minerals are frequently bounded by their own crystal faces (idioblasts). Larger crystals are often packed with small inclusions of other minerals exhibiting the so-called sieve structure (poikilitic or diablastic structure).

Granoblastic refers to a nondirected rock fabric with minerals forming grains without any preferred shape or dimensional orientation (Fig. 1*c*). Lepidoblastic (Fig. 1*d*), nematoblastic (Fig. 1*e*), and fibroblastic refer to rocks of scaly, rodlike, and fibrous minerals, respectively.

The metamorphic minerals may be arranged in an idioblastic series (crystalloblastic series) in their order of decreasing force of crystallization as follows: (1) sphene, rutile, garnet, tourmaline, staurolite, kyanite; (2) epidote, zoisite; (3) pyroxene, hornblende; (4) ferromagnesite, dolomite, albite; (5) muscovite, biotite, chlorite; (6) calcite; (7) quartz, plagioclase; and (8) orthoclase, microcline. Crystals of any of the listed minerals tend to assume idioblastic outlines at surfaces of contact with simultaneously developed crystals of all minerals of lower position in the series.

Preferred orientation. Certain minerals have a tendency to assume parallel or partially parallel

Microstructures of uranium-carbon alloys: (a) 4.46% C; (b) 4.86% C; (c) 5.75% C; (d) 7.98% C; (e) 9.08% C; all ×50. Samples were prepared by arc melting uranium and carbon in an argon atmosphere on a copper-cooled hearth. The ground and polished surfaces were chemically stained to produce a thin oxide film, the thickness and resulting color of which is dependent upon the crystallographic orientation of the substrata. The color indicates the grain size and amount and location of the different phases. (USAEC-Union Carbide Corp. Oak Ridge Natl. Lab.)



(a)

(b)



(c)

(d)







Fig. 1. Photomicrographs of typical structures of brass (70% Cu, 30% Zn). (a) Annealed (grains with twin bands). (b) Reduced 40% by cold rolling (distorted grains) (c) Stress-corroded specimen (note crack). All 100X. (W. R Johnson)

magnifications are attained with electron microscopy. *See* MICROSCOPE, ELECTRON.

Metallography serves research and industrial practice. It is used in the investigation of many phenomena of physical metallurgy, including the grain structures of metals, their deformation mechanisms, and most important, their phase constitution. Industry uses metallographic methods for the control of production processes, especially heat treatments. These methods are also useful in the analysis of the causes of failure of metallic objects.

Applications. The photomicrographs in Figs. 1, 2, and 3 illustrate some of the typical applications of metallography. Typical phenomena include grain structure, size, and shape (Figs. 1*a* and 2*a*); substructure, subboundaries (veining); crystal orientation, etch pits (Fig. 3*c*); inhomogeneities, dendritic microsegregation; microporosity; microconstituents, phases or aggregates of phases forming microconstituents (Fig. 2*b* and *c*); phase changes, hot; polymorphic transformations; precipitation, size, shape, and arrangement of precipitate (Widmannstätten structures); effects of deformation, strain markings, distorted grains (Fig. 1*b*), mechanical twins, and microcracks; structure of steel and other industrial metals; heat-treating defects, burned and overheated steel, decarburized layers; surface layers, carburized cases, electroplate; and corrosion, stress corrosion (Fig. 1*c*), and dezincification.

Preparation of specimens. The selection of representative specimens is essential. If a part, such as a forged shaft, has directional properties, transverse and longitudinal sections are prepared. Sheet, wire, and other small specimens are mounted in fixtures or plastic mounts.

A specimen is first ground on a series of emery papers of decreasing grit size or on laps. It is then polished on one or more cloth-covered wheels with an abrasive such as aluminum oxide, magnesium oxide, or diamond dust. These operations are designed to render the specimen surface scratch-free and mirrorlike by the progressive removal of surface irregularities, but polishing, even when properly carried out, produces a thin layer of distorted metal

Electrolytic polishing is an alternative to mechanical polishing. Electrolytic polishing consists of controlled anodic dissolution; thus it does not distort the surface layer of the metal, and is therefore particularly suitable for soft metals. However, inclusions and compounds present in a specimen may react with the electrolyte. Once the operating conditions have been established, electrolytic polishing is simple, and it prepares a surface which



Fig. 2. Photomicrographs of typical structures of iron and steel. (a) High-purity iron, 100X. (b) Steel containing 0.85% C, slowly cooled (pearlite structure composed of light ferrite and dark cementite), 500X. (c) Steel containing 0.80% C, quenched and tempered (needles of tempered martensite), 500X. (d) Steel containing 1.10% C, cooled at moderate rate (a Widmannstätten structure of white cementite plates in a matrix of fine pearlite), 200X. (W. R. Johnson)

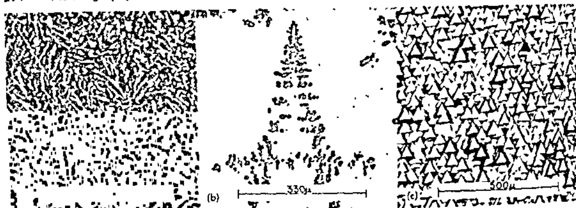


Fig 3. Photomicrographs of typical structures. (a) Eutectic in an alloy of lead and antimony containing 11% Sb, 89% Pb (mixture of two phases solidified together), 750X (b) A large dendrite (treelike structure) of copper oxide in a matrix of copper–copper oxide eutectic (branches of dark particles grown from a single trunk), 150X. (c) Etch pits in germanium (identical orientation of pits indicating single grain), 100X. (*W. R. Johnson*)

may be superior to a mechanically polished surface.

Polished specimens reveal only some structural features, such as microporosity and inclusions. Etching with an appropriate chemical is necessary to bring out the microstructures. Etching also removes the distorted metal from mechanically polished specimens.

Etching reagents are acids, alkalies, or complex substances, generally in dilute solution in water, alcohol, or glycerin. Most reagents act by dissolution, a few by selective deposition of reaction products (staining). In a single-phase alloy, etching is effective because it attacks different parts of the structure differently. Grain-boundary regions dissolve in preference to the body of the grains; the resulting grooves are observed as a dark network, as shown schematically in Fig. 4*a* and illustrated in the photomicrograph in Fig. 2*a*. Adjoining grains may develop facets of different orientation, which reflect differing amounts of light as shown in Fig. 4*b* and illustrated in Fig. 1*a*. In multiphase alloys, the phases are attacked selectively; Figs. 2*b*, 3*a*, and 3*b* show examples of this.

Etching may also be carried out in an electrolyte

men is bombarded by ions of an inert gas, is effective with some metals which resist etching by other methods.

Metallographic microscopy. The microscopic examination of metallographic specimens depends on the differences in the amount and quality of light reflected by different structural constituents. *See* MICROSCOPE, OPTICAL.

In the standard metallographic microscope, a beam of light normal to the surface illuminates the specimen. The objective of the microscope serves as a condensing system for the incident beam and as an image-forming system for the light reflected by the specimen. This is known as bright-field illumination. Oblique illumination by an off-center beam improves contrast by casting shadows. Conical-stop illumination achieves the same effect. In dark-field illumination, a hollow beam of light converges on the specimen; an ideal mirror appears dark, while rough surfaces appear bright.

Polarized light permits identification of some constituents, especially nonmetallic inclusions. It also may bring out contrast between adjoining grains, particularly in anisotropic metals. Ultraviolet light improves resolution, but has found little use in metallography. In phase-contrast microscopy, differences in levels of the specimen surface appear as differences in brightness. Phase contrast is also sensitive to the nature of the re

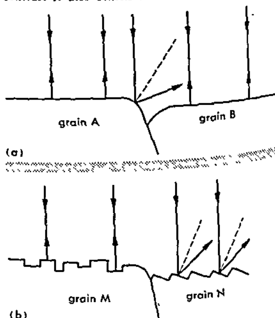


Fig. 4. The reflection of light by etched metal surfaces (schematic). Both diagrams normal to specimen surface (a) Preferentially etched grain boundary (dark in image). (b) Etched facets on two adjoining grains (differences in light reflected).

flecting surfaces. *Fine* structural details (veining, slip lines, and precipitates) have been observed by phase contrast when ordinary methods failed. Interference microscopy permits quantitative investigation of surface contours.

Observation at elevated temperatures allows processes such as recrystallization or precipitation to be followed as they occur. The microscope stage must accommodate a heating device, and most specimens must be protected from oxidation by a protective atmosphere or vacuum. Also, the objective must be kept cool. Their long working distance makes reflecting objectives particularly suitable for this application (*see* MICROSCOPE, REFLECTING). In high-temperature metallography, thermal etching reveals the structural features.

In microautoradiography, radiations emitted by radioactive tracers are recorded on photographic emulsions, which are then investigated microscopically. Conventional metallography of highly radioactive metals calls for equipment operated entirely by remote control. Specimens of toxic metals are prepared in airtight enclosures (glove boxes).

Photomicrographs of metallographic specimens are often taken. For this purpose, special microscopes known as metallographs have been developed.

Quantitative metallography. In quantitative metallography, provided the specimen is representative, the problem is one of translating measurements on a two-dimensional section into quantities for the three-dimensional structure existing in the metal.

Methods for measuring grain size consist of matching a microsection with published standards, counting the number of grains intersected by lines of known length or counting the number of grains enclosed in a known area. The particle sizes of a phase dispersed in another phase can also be determined; the solution of this problem depends upon the shape of the particles and their size distribution.

The relative amounts of each of several phases present in a multiphase alloy are often of interest. The volume fractions are equal to the fractions of the areas occupied by each of the phases in a random section; the areas are measured directly or found from superposed random lines or points (lineal analysis and point counting). The total area of internal boundaries (grain boundaries) in a given volume is related to the number of intersections of a random line of given length with the traces of the boundaries in a random section.

Factors governing microstructures. The factors governing the formation of microstructures are largely understood. For example, on approaching equilibrium at elevated temperatures, grains of a single-phase alloy tend to join along a common edge in groups of three, at equal angles of 120°, because this minimizes grain boundary area and hence the interfacial energy. Annealing twins (Fig. 1*a*) originate in growth accidents during recrystallization. Columnar grains, microsegregation, and the structures of eutectics can be related to the solidification processes by which they form.

Reactions in solid metals also produce characteristic microstructures. The lamellae of eutectoids such as pearlite in steel (Fig. 2*b*) result from the nature of a decomposition reaction. The precipitation of a phase from a supersaturated matrix often produces a regular arrangement of the precipitate particles (a Widmannstätten structure, Fig. 2*d*) because they have a definite orientation relation to the crystal structure of the matrix. Microstructures thus result from the interplay of various factors, such as rates of growth, interfacial energy, and crystallographic relations. *See* ALLOY; CRYSTAL STRUCTURE; HEAT-TREATMENT (METALS AND ALLOYS); STEEL. [M.B.]

Bibliography: R. M. Brick and A. Phillips, *Structure and Properties of Alloys*, 2d ed., 1949; G. L. Kehl, *Principles of Metallographic Laboratory Practice*, 3d ed., 1949

Metalloid

An element which exhibits the external characteristics of a metal but behaves chemically both as a metal and as a nonmetal. Arsenic and antimony, for example, are hard crystalline solids that are definitely metallic in appearance. They may, however, undergo reactions that are characteristic of both metals and nonmetals. Certain of their oxides dissolve in either acids or bases, and are said to be amphoteric in character because they behave either as a base or an acid. Many elements form compounds that are amphoteric. However, only when this dualistic chemical behavior is very marked and the external appearance metallic is the element commonly called a metalloid. *See* METAL; NONMETAL. [F.J.J.]

Metallurgy

The extraction from ores, refining, and preparation for use of the various metals. For purposes of discussion, teaching, and specialization, the field of metallurgy has been divided in various ways. Partly for historical and economic reasons, but chiefly because of wide differences in technological details, nonferrous metallurgy is distinguished from ferrous metallurgy. Each of these may be subdivided into process or extractive metallurgy, fabrication metallurgy, and physical metallurgy. The extractive metallurgist is concerned with producing metals from ores or concentrates; he uses the methods of pyro-, electro-, or hydrometallurgy. The fabrication metallurgist converts the raw metals into alloys and into various shapes such as rails, sheet, wire, and automobile bodies. The physical metallurgist studies the structure and properties of metals and alloys for purposes of industrial control or scientific investigation. The subdivisions mentioned above by no means exhaust the list of specialties in the field. Practically every one of these subdivisions may appear under each of the major and many of the minor metals. The rolling of brass is not the same as rolling of steel, although both operations use rolling mills. Steel refining, nickel re-

fining, and copper matte production all use reverberatory furnaces but have few other similarities. The physical metallurgist who is familiar with aluminum and its alloys may know little about steel or magnesium. The field of metallurgy is far from devoid of general principles but their applications require intimate knowledge of the characteristics of the individual metals. *See* ALLOY; ELECTROMETALLURGY; ELECTROPLATING OF METALS; HYDROMETALLURGY; IRON (EXTRACTION FROM ORE); METAL COATINGS; METAL FORMING; METAL INSPECTION, MAGNETIC; METAL INSPECTION, ULTRASONIC; METALLOGRAPHY; ORE DRESSING; PYROMETALLURGY, NONFERROUS; STEEL MANUFACTURE; see also articles on individual metals. [E.R.JE.]

Metamerism

The serial repetition of parts along the longitudinal body axis in bilaterally symmetrical animals. It is also called segmentation. The successive subdivisions are called segments, somites, or metameres. Common examples are the muscles and spinal nerves in the human body and in the body and tail of many mammals, snakes and lizards, salamanders, and fishes. If repeated parts are essentially alike, as along the body of a snake or earthworm, the segmentation is homonomous. When unlike, as in somites of the crayfish or insects, it is termed heteronomous. In the heads of arthropods, several primitive embryonic somites become fused in the adults. [T.I.S.]

Metamerism, embryonic

The repetitive form [illegible] and certain jellyfish. True metamerism occurs in annelid worms, arthropods, vertebrates and other chordates. It originates either from a bilateral series of coelomic pouches which form the segmental muscles, kidneys, and body cavities of lower forms, or from mesoblastic somites which form the skeletal and muscular segments of vertebrates. Repetitive features of the nervous system are acquired secondarily through the influences of mesodermal metameres upon adjoining ectodermal tissues. *See* COELOM; EMBRYOLOGY; NEURULATION. [H.L.H.]

Metamorphic rocks

One of the three major groups of rocks that make up the crust of the earth. The other two groups are igneous rocks and sedimentary rocks. Metamorphic rocks are pre-existing rock masses in which new minerals, or textures, or structures are formed at higher temperatures and greater pressures than those normally present at the earth's surface. *See* IGNEOUS ROCKS; SEDIMENTARY ROCKS.

Two groups of metamorphic rocks may be distinguished: cataclastic rocks, formed by the operation of purely mechanical forces; and recrystallized rocks, or the metamorphic rocks properly so called, formed under the influence of metamorphic pressures and temperatures.

Cataclastic rocks are mechanically sheared and crushed. They represent products of dynamometamorphism, or kinetic metamorphism (*see* METAMORPHISM). Chemical and mineralogical changes generally are negligible. The rocks are characterized by their minute mineral grain size. Each mineral grain is broken up along the edges and is surrounded by a corona of debris or strewn fragments (mortar structure, Fig. 1*a*). During the early stages of this alteration process the metamorphosed product is known as flaser rock (*Fig. 1b*). Eventually the original mineral grains are entirely gone, as in the mylonites. When seen through the microscope the comminuted particles consist of a mixture of finely powdered quartz, feldspar, and other minerals with an incipient recrystallization of sericite or chlorite. Pseudotachylite is an extreme end product of this crushing process. *See* FLASER ROCK; MYLONITE.

Structural relations. Metamorphic rocks, properly so called, are recrystallized rocks. The laws of recrystallization are not the same as those of simple crystallization from a liquid, because the crystals can develop freely in a liquid, but during recrystallization the new crystals are encumbered in their growth by the old minerals. Consequently the structures which develop in metamorphic rocks are distinctive and of great importance, because in many ways they reflect the physicochemical environment of recrystallization and thereby the genesis and history of the metamorphic rock.

Crystalloblastic structure. A crystalloblast is a crystal that has grown during the metamorphism of a rock. The majority of the minerals in metamorphic rocks are irregular in outline (xenoblasts), but some minerals are frequently bounded by their own crystal faces (idioblasts). Larger crystals are often packed with small inclusions of other minerals exhibiting the so-called sieve structure (poikilitic or diablastic structure).

Granoblastic refers to a nondirected rock fabric with minerals forming grains without any preferred shape or dimensional orientation (Fig. 1*c*). Lepidoblastic (Fig. 1*d*), nematoblastic (Fig. 1*e*), and fibroblastic refer to rocks of scaly, rodlike, and fibrous minerals, respectively.

The metamorphic minerals may be arranged in an idioblastic series (crystalloblastic series) in their order of decreasing force of crystallization as follows: (1) sphene, rutile, garnet, tourmaline, staurolite, kyanite; (2) epidote, zoisite; (3) pyroxene, hornblende; (4) ferromagnesite, dolomite, albite; (5) muscovite, biotite, chlorite; (6) calcite; (7) quartz, plagioclase; and (8) orthoclase, microcline. Crystals of any of the listed minerals tend to assume idioblastic outlines at surfaces of contact with simultaneously developed crystals of all minerals of lower position in the series.

Preferred orientation. Certain minerals have a tendency to assume parallel or partially parallel

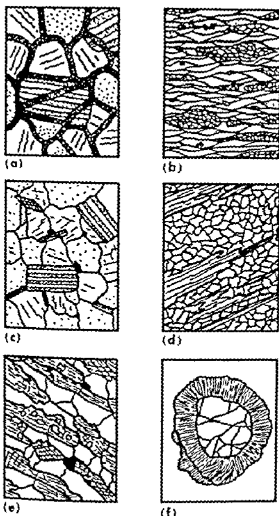


Fig. 1. Some fabrics of metamorphic rocks as seen under the microscope. (a) Mortar fabric, (b) flaser, or mylonitic, fabric; (c) granoblastic fabric (from E. E. Wahlstrom, *Petrographic Mineralogy*, Wiley, 1955). (d) lepidoblastic fabric; (e) nematoblastic fabric, (f) porphyroblast with reaction rim (from T. F. W. Barth, *Theoretical Petrology*, Wiley, 1952).

crystallographic orientation. The shape and spatial arrangement of minerals, such as mica, hornblende, or augite, show a definite relation to the foliation in the schist or gneiss; that is, both foliation and fissility of a metamorphic rock are directly related to the preferred position assumed by the so-called schist-forming minerals, such as mica, hornblende, and chlorite. *See* GNEISS; SCHIST.

Students of structural petrology distinguish between preferred orientation of inequidimensional grains according to their external crystal form, and preferred orientation of equidimensional grains according to their internal or atomic structure. *See* PETROFABRIC ANALYSIS; STRUCTURAL PETROLOGY.

A special microscope technique (universal stage technique) is necessary in most cases to demonstrate in detail the preferred orientation of the mineral grains according to their atomic structure. *See* PETROGRAPHY.

Relic structures. Mineral relics often indicate the temperature and pressure that obtained in the preexisting rock. If a mineral is stable in the earlier rock and is also stable in the later rock, it will, unless stress action sets in, be preserved in its original form and present a stable relic. Common examples are quartz in various rocks and feldspar in amphibolites. However, when a mineral or a definite association of minerals becomes unstable, it may still escape alteration and appear as an unstable relic. These relics are proterogenic, that is, representative of an earlier, premetamorphic rock, or of an earlier stage of the metamorphism. Hysterogenic products are of later date, and are formed in consequence of changed conditions after the formation of the chief metamorphic minerals.

A common phenomenon, fairly illustrative of the tendency toward equilibria, is the formation of armors or reaction rims around minerals (Fig. 1*f*) which have become unstable in their association but have not been brought beyond their fields of existence in general (the armored relics). Thereby the associations of minerals in actual contact with one another become stable. If, however, the constituent minerals of a rock containing armored relics are named without noting this phenomenon, it may be taken as an unstable association. *See* PORPHYROBLAST.

Structure relics are perhaps of still more impor- [illegible] ...icating the nature of the preex- [illegible] ...etamorphic deform... [illegible] compared to the readi... [illegible] ...en... [illegible] ...ch- ments used for the second time after original writing was nearly erased. Every trace of original structure is important in attempting to reconstruct the history of the rock and in analyzing the causes of its metamorphism.

In sedimentary rocks the most important structure is bedding (stratification or layering) which originally was approximately horizontal. In metamorphic rocks deformed by folding, faulting, or other dislocations, the sum of all deformations can

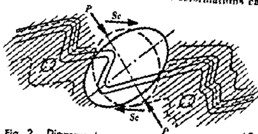


Fig. 2. Diagram showing the general relationship between deformation folding and slaty cleavage caused by pressure *PP* or the couple *ScSc*. Heavy black lines denote the original bedding deformed into folds. Thin lines indicate slaty cleavage which may grow into schistosity (false bedding). (After G. Wilson, Proc. Geol. Assoc., 1946)

be referred to the original horizontal plane, and the deformations can be analyzed.

Fissility and schistosity. One of the earliest secondary structures to develop in sediments of low metamorphic grade is that of slaty cleavage (also referred to as flow cleavage or fissility), which grades into schistosity which is different from fracture cleavage, or strain-slip cleavage. Slaty cleavage is developed normal to the direction of greatest shortening of the rock mass, and cuts the origi- [illegible] cleavage develops in response to the stresses imposed on the rock system as a whole because of the differential resistance of the several layers. Consequently, folding and slaty cleavage have a common parentage, as illustrated in Fig. 2

In the rock series slate-phyllite-schist the slaty cleavage will grade into schistosity. It is a chemical and recrystallization phenomenon as well as a mechanical one, and the directions of the schistosity become the main avenues of chemical transport. *See* CLEAVAGE, ROCK; SCHISTOSITY, ROCK; SLATE.

Contact-metamorphic rocks. Igneous magma at high temperature may penetrate into sedimentary rocks, it may reach the surface, or it may solidify in the form of intrusive bodies (plutons). Heat from such bodies spreads into the surrounding sediments, and because the mineral assemblages of the sediments are adjusted to low temperatures, the heating-up will result in a mineralogical and textural reconstruction known as contact metamorphism

The width of the thermal aureole of contact metamorphism surrounding igneous bodies varies from almost complete absence in the case of small intrusions (basalt dikes or diabase sills) to several kilometers in the case of large bodies. *See* AUREOLE, CONTACT.

The effects produced do not depend only upon the size of the intrusive. Other factors are amount of cover and the closure of the system, composition and texture of the country rock, and the abundance of gaseous and hydrothermal magmatic emanations. The heat conductivity of rocks is so low that gases and vaporous emanations become chiefly responsible for the transportation and transfer of heat into the country rock. *See* ORE DEPOSITS, GEOCHEMISTRY OF.

Alteration of stratified rocks. Stratified rocks are altered in the contact zone to what is commonly called hornfels or hornstone. They are hardened, often flinty rocks, usable for road material and so fine-grained that the mineral components can be discerned only with the microscope. Hornfels used to be regarded as "silicified" sediments. However Kjerulf, in the latter half of the nineteenth century, analyzed sedimentary shale and "silicified" shale of the Oslo region and found that, chemically, they were identical (except for water and carbon dioxide content). Then geologists realized that the "hardening" of the shale took place without appreciable change in the chemical composition. Kjerulf summarized his results by saying that the composition of a hornfels depended on the original sediment (the shale) and was independent of the kind of adjacent igneous rock.

Later H. Rosenbusch arrived at the same conclusion and pronounced that no chemical alterations accompany the formation of hornfelses except for the removal of fugitive constituents. The Kjerulf Rosenbusch rule is useful but needs modification, because chemical changes may ensue from hydrothermal and pneumatolytic action.

The next problem, then, is to see how the mineral assemblages of the hornfelses depend upon the chemical composition of the original sediments. The chief types of sedimentary rocks are sandstone (sand), shale (clay), and limestone. Among the varieties of hornfelses which may develop from different mixtures of these components, the continuous series from shale to limestone is the most interesting.

Most shales contain some iron- and magnesia bearing constituents in addition to feldspar and clay minerals. Quartz (SiO_2) is always admixed [illegible] is often present in the [illegible] rals. Other [illegible] tituents are [illegible] so constituents are grouped together to define a system of three components, alumina, lime, and ferromagnesia.

By applying the mineralogical phase rule, which states that the number of stable minerals in a rock shall not be larger than the number of components, it follows that, except for quartz and some alkali-bearing minerals listed below, no more than three additional minerals should occur in any one (variety) of these hornfelses. Observations have verified this. Thus from alumina, lime, and ferromagnesia, seven minerals will form that are stable under the conditions of contact metamorphism: andalusite, Al_2SiO_5; cordierite, $Mg_2Al_4Si_5O_{18}$; anorthite, $CaAl_2Si_2O_8$; hypersthene, $(Mg,Fe)SiO_3$; diopside, $Ca(Mg,Fe)Si_2O_6$; grossularite, $Ca_3Al_2Si_3O_{12}$; wollastonite, $CaSiO_3$. Only three (or fewer) of these minerals can occur together. In this way different mineral combinations develop, each combination (plus quartz and an alkali-bearing mineral) representing a natural hornfels. There are 10 such combinations, corresponding to hornfelses of classes 1–10 of V. M. Goldschmidt's terminology. *See* HORNFELS.

Variations from the above scheme are easily explained. Usually enough water and potash are present to produce mica; muscovite may form instead of, or together with, andalusite, and in the hornfelses of classes 4 and 5, biotite is usually present inducing a characteristic chocolate-brown color into the rocks. In hornfelses of class 10 some lime-rich hydrous silicates may develop, for example

vesuvianite (idocrase). The presence of ferric iron may produce andradite ($Ca_3Fe_2Si_3O_{12}$), a yellow to dark green garnet which will form mixed crystals with grossularite.

Pneumatolysis and metasomatism. Other factors of importance in contact metamorphism are chemical changes that ensue from pneumatolytic and hydrothermal action. These changes are brought about by the magmatic gases and high temperatures that accompany igneous intrusions. The surrounding rocks are deeply penetrated not only by the heat but also by water and other volatile compounds. Because chemical alterations take place in this so-called pneumatolytic or hydrothermal contact metamorphism, the Kjerulf-Rosenbusch rule is not applicable. Hornfelses may be affected several feet from the contact. Large rock masses are sometimes affected. *See* METASOMATISM; PNEUMATOLYSIS.

The primary magmatic gases are acid and in consequence show high reactivity. If the contact rock is basic, especially limestone, the acid gases will react effectively with it. Limestone acts as a filter, capturing the escaping gases. As a result a great variety of reaction minerals are formed. The corresponding rocks are known as skarns If the reaction rocks are limestones composed of lime silicates, the reaction minerals are mainly garnet and pyroxene, often accompanied by fluorite and phlogopite. Sulfides of iron, zinc, lead, or copper may be present and in some occurrences magnetite is formed. *See* SKARN.

Summary. Metamorphism in deep-seated rocks is very common, and the products (disregarding the pneumatolytic action) vary regularly in accordance with the chemical compositions of the preexisting contact rock. Another factor of equal importance is the variation in temperature as influenced by the nature of the intruding rock and the distance from the contact. Thus it is possible to distinguish between an inner and an outer contact zone. The zones grade into each other by imperceptible transitions, but the mineral associations in the typical inner contact zone, the only zone considered so far, are markedly different from the associations in the outer contact zones.

These problems involve a consideration of the general relationships between the minerals and mineral associations, on the one hand, and temperature and pressure, on the other. They are discussed further in connection with the facies principle and the general process of regional metamorphism However, it is important to realize that contact metamorphism, although it appears to be well defined and seems to stand out as an isolated natural phenomenon, is complex and variegated and passes by gradual transitions into other kinds of metamorphism. Geologically, contact metamorphism should be considered in connection with, and as a part of, the general system of rock metamorphism and metasomatism.

Regional metamorphic rocks. Crystalline schists, gneisses, and migmatites are typical products of regional metamorphism and mountain building. If sediments accumulate in a slowly subsiding geosynclinal basin, they are subject to down-warping and deep burial and thus to gradually increasing temperature and pressure. They become sheared and deformed, and a general recrystallization results. However, subsidence into deeper parts of the crust is not the only reason for increasing temperature It is not known what happens at the deeper levels of a live geosyncline, but obviously heat from the interior of the earth is introduced regionally and locally, partly associated with magmas, partly in the form of "emanations" following certain main avenues, determined by a variety of factors. *See* EARTH (HEAT FLOW). From this milieu rose the lofty mountain ranges of the world, with their altered beds of thick sediments of the shallow seas and intercallations of tuffs, lavas, and intrusives (plutons), all thrown into enormous series of folds and elevated to thousands of meters. Thus were born the crystalline schists with their variants of gneisses and migmatites. *See* OROGENY.

According to Rosenbusch, orthogneisses develop from preexisting igneous rocks, paragneisses from sediments. In addition, there are mixed rocks (migmatites) in which extended recrystallization, replacement, and metasomatism have resulted in a complete chemical reconstitution.

A. Michel-Lévy (1888) distinguished three main étages in the formation of the crystalline schists; F. Becke and U. Grubenmann (1910) demonstrated that the same original material may produce radically different metamorphic rocks according to the effective temperature and pressure during the metamorphism. Grubenmann distinguished three successive depth zones, epizone, mesozone, and katazone, corresponding to three consecutive steps of progressive metamorphism. In eroded mountain ranges, rocks of the katazone are, generally speaking, encountered in the central parts; toward the marginal parts are found rocks of the mesozone and epizone.

It is of paramount importance to obtain better information about the temperature-pressure conditions of the recrystallization and thus to show the relation between the chemical and mineralogical composition of all varieties of rocks. A large-scale attempt in this direction was the development of the facies classification of rocks.

Mineral facies As defined by P. Eskola (1921), a mineral facies "comprises all the rocks that have originated under temperature and pressure conditions so similar that a definite chemical composition has resulted in the same set of minerals, quite regardless of their mode of crystallization, whether from magma or aqueous solution or gas, and whether by direct crystallization from solution . . . or by gradual change of earlier minerals . . ." To find out which mineral associations were characteristic of high temperature or of low temperature, and to determine which associations combined with high pressure and with low pressure, Eskola studied the mineral associations in the rocks.

It has long been known that in an area of progressive metamorphism each successive stage, or each new zone of metamorphism, is reflected in the appearance of characteristic rock types (G. Barrow, 1893). Rocks within the same zone may be called isofacial, or isograde as proposed by C. E. Tilley (1924) who, furthermore, proposed the term "isograd" for a line of similar degree of metamorphism.

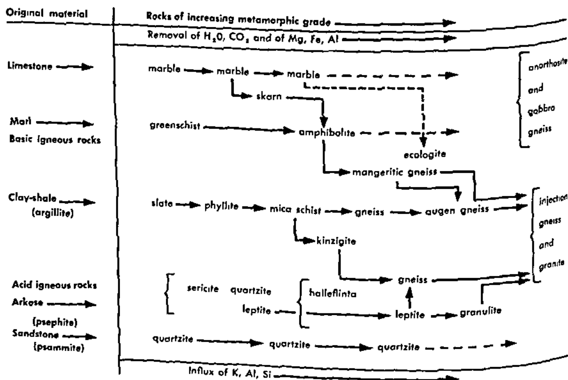
In going from an area of unmetamorphosed sedimentary rocks into an area of progressively more highly metamorphic rocks, new minerals appear in orderly succession. Thus, in a series of argillaceous rocks subjected to progressive metamorphism, the first index mineral to appear is usually chlorite, followed successively by biotite, garnet (almandite), and sillimanite. A line can be drawn on the map indicating where biotite first appears. This line is the biotite isograd. The less metamorphosed argillites on one side of this line lack biotite, whereas the more metamorphosed rocks on the other side contain biotite. An isograd can be drawn for each mineral. Actually the isograds are surfaces, and the lines drawn on the map are the intersections of these surfaces with the surface of the earth.

Further work along these lines resulted in the conclusion that it was possible to single out a well-defined series of mineral facies. Sedimentary rocks of the lowest metamorphic grade recrystallized to give rocks of the greenschist facies—chlorite and epidote being characteristic minerals. A higher degree of metamorphism produced the epidote-amphibolite facies and a still higher degree the true amphibolite facies in which hornblende and plagioclase mainly take the place of chlorite and epidote. Representative of the highest regional metamorphic grade is the granulite facies in which most of the stable minerals are water-free, such as pyroxenes and garnets. Any sedimentary unit will recrystallize according to the rules of the several mineral facies, the complete sequence of events being a progressive change of the sediment by deformation, recrystallization, and alteration in the successive stages, greenschist facies → epidote amphibolite facies → amphibolite facies → granulite facies.

During regional metamorphism a stationary temperature gradient is supposed to be established in the mountain masses. Usually, the outer parts of a geosynclinal region are less affected, and in the ideal case the marginal parts contain unmetamorphosed sediments, clay, sand, and limestone, which gradually change into metamorphic rocks of successively higher facies as they extend into the central and deeper parts (depth zones). *See* GEOSYNCLINE.

The accompanying table summarizes the metamorphic series of rocks that develop from the several types of common sediments and usually converge toward a granitic composition regardless of the nature of the original material. The chemical changes in the progressive series are characterized by increase in alkali metals and silica, and decrease in magnesium, calcium, and iron (see Fig. 3).

In the accompanying table original basic igneous rocks (gabbros, basalts) show a composition related to that of marl and yield analogous metamorphic products. Not listed are ultrabasites (peridotites, and others) which by metamorphism become serpentine, chlorite or talc schist, soapstone, hornblende schist, pyroxene or olivine masses. Original acid igneous rocks (granite, diorite, rhyolite) show a composition related to that of arkose and



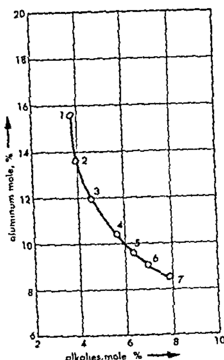


Fig 3. Progressive chemical changes in metamorphic sediments, with increasing regional metamorphism 1, Sedimentary shale (average of 27); 2, sericite schist (av. 25); 3, two-mica schist (av. 37); 4, two-mica gneiss (av. 51); 5, biotite gneiss and granitoids (av 64); 6, granite (av. 74); 7, granulite (av. 24). (After P. Lapadu-Hargues, Compt. rend. soc. géol., France, 2:11–13, 1945)

yield analogous products. Leptite is primary fine-grained, usually showing tufaceous or blastoporphyric relict structures; or it is derived from argillaceous sediments. Hälleflintas are dense rocks of conchoidal fracture, genetically related to leptites. Kinzigites, characterized by containing aluminum silicates, usually also rich in magnesia, are metasomatic gneisses, but probably argillites also enter into their constitution. Granulite used to be a neutral name designating medium- to fine-grained gneisses. In Germany and Fennoscandia, granulite means a garnet gneiss with but little mica and a characteristic structure (granulite facies).

schists, all of which are green, hence the name. Other common rocks in this group are serpentinites, talc schists, phyllites, and muscovite (sericite) schists. Plagioclase is not stable in greenschists but breaks up into epidote and albite. Glaucophane schists are rare, and probably represent greenschists formed under high stress.

Rocks of the epidote-amphibolite facies recrystallized in a somewhat higher temperature range. Amphibolites with epidote and either albite or oligoclase are typical. Mica schists (garnet), biotite schists, staurolite schists, and kyanite schists are common. Cordierite-antophyllite (gedrite) ... and chloritoid schists also occur. Sodic pla-

gioclase are typical and often carry quartz or biotite (or both) or garnet. Sillimanite-muscovite schists (gneisses) are common.

Rocks of the granulite facies have their greatest extension in old Precambrian areas, but are also found in younger deeply eroded mountain chains. Hypersthene (garnet) plagioclase gneisses, usually ... occasionally with olivine or spinel

perature range is probably from about 500–700°C.

Experimental work. The experiments by H. S. Yoder (1952) in the system $MgO—Al_2O_3—SiO_2—H_2O$ indicated that at approximately 600°C and 1200 atm it is possible to have different mineral assemblages suggestive of every one of the now accepted metamorphic facies in stable equilibrium. These different mineral assemblages (artificial facies) observed by Yoder are the result of differences in the water content, and are not related to variation in temperature and pressure.

As an example, the mineral clinochlore, corresponding to one of the most common rock-making chlorites, shows an upper limit of stability, either alone or in association with talc, of 680°C. This is, indeed, a superhigh temperature for any kind of metamorphism, almost a magmatic temperature. But according to the tenets of the mineral facies, chlorite is strictly limited to the greenschist facies, about 200°C.

Temperature and mineral facies. Although Yoder has proved that there is no absolute relation between temperature and facies, it appears likely that, to the field geologist, and to the laboratory man as well, the established facies classification will still remain the best system of classification of

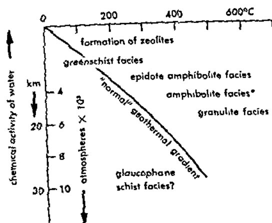


Fig 4. The mineral facies of regional metamorphic rocks in their normal relation to temperature, pressure and chemical activity of water.

metamorphic rocks; and in a majority of cases the facies will indicate the temperature-pressure conditions under which the several rocks recrystallized. Generally speaking, there is a regular relation between the chemical activity of water and the facies.

Water content and mineral facies. The role of water in metamorphism is determined by at least four variable, geologically related parameters, rock pressure, temperature, water pressure, and the amount of water present. During a normal progressive regional metamorphism, rock pressure and temperature are interdependent, and the amount of water and the pressure of water are related to the encasing sediments and to the degree of metamorphism in such a way that, generally speaking, the low-grade metamorphic facies are characterized by the presence of an excess of water, the medium-grade by some deficiency in water, and the high-grade metamorphic facies by virtual absence of water.

In the usual diagrammatic illustration of the mineral facies of rocks, temperature and pressure *(depth)* are taken as coordinates; in regional-metamorphic rocks a third, dependent coordinate may be added: the activity of water running parallel to, *but in the* opposite direction of, the pressure (Fig 4).

Summary. It is still impossible to define with any *confidence the* temperature-pressure fields of the various metamorphic facies. Figure 4 summarizes only tentative conclusions to the question of the possible normal *relations* between temperature and mineral facies of regional-metamorphic rocks. The mineral facies of a rock is not a single-valued function of temperature (and *pressure*). *At* the present state of knowledge it is impossible to infer with certainty the temp*erature at which* the metamorphism took place using the mineral facies concept. See AMPHIBOLITE; ECLOGITE; EPIDIORITE; EPIDOSITE; GRANULITE; GREISEN; MARBLE; MICA SCHIST; MIGMATITE; PHYLLITE; PHYLLONITE; QUARTZITE; SANIDINITE; SCAPOLITE; SERPENTINITE; SOAPSTONE.

[T.F.W.B.]

Bibliography: T. F. W. Barth, *Theoretical Petrology*, 1952; W. H. Bucher, Fossils in metamorphic rocks, *Geol. Soc. Am., Bull.* 64:275–300, 1953; T. F. W. Barth, C. W. Correns, P. Eskola, *Die Entstehung der Gesteine*, 1939; W. S. Fyfe, F. J. Turner, and J. Verhoogen, *Metamorphic Reactions and Metamorphic Facies*, Geol. Soc. Am. Mem. 73, 1958; A. Harker, *Metamorphism*, 1932; H. Ramberg, *Origin of Metamorphic and Metasomatic Rocks*, 1952; F. J. Turner, *Mineralogical and Structural Evolution of Metamorphic Rocks*, Geol. Soc. Am. Mem. 30, 1948; F. J. Turner and J. Verhoogen, *Igneous and Metamorphic Petrology*, 1951; H. S. Yoder, The $MgO—Al_2O_3—SiO_2—H_2O$ system and the related metamorphic facies, *Am. J. Science*, Bowen vol. (2):569–627, 1952.

Metamorphism

The alterations and transformations in preexisting rock masses effected by temperature and pressure, but excluding changes produced by weathering and sedimentation. The changes may include the production of new minerals, structures, or textures, or all three. They give a distinctive new character to the rock as a whole, but they do not involve the loss of individuality of a rock mass, such as changes brought about by fusion. Quantitatively, the metamorphic rocks, including gneisses and migmatites, are the most important group of rocks in the crust of the continents.

James Hutton (1726–97) of Edinburgh was the first to advance the opinion that some of these rocks were ordinary sedimentary strata which were sub-[illegible] heat as to [illegible] term metamorphic for the altered strata.

Both igneous rocks and sediments are subject to alteration. The former are fusion products of processes which operated at high temperatures whereas sediments are deposits which accumulated under conditions of low temperature. Once formed, the rocks of both classes respond to moderately elevated temperatures and changes in pressure, or shearing stresses, by reaction and recrystallization.

Metamorphic processes cannot be directly observed in natural rocks, nor have they been adequately investigated by experiments. For this reason the mechanism and thermodynamic relations of metamorphic processes are not completely understood. The study of transitional rock series ranging from unaltered rock products to completely altered, or recrystallized, rocks provides the data on which the principles of metamorphic geology are formulated.

Metamorphism versus metasomatism. The alteration of rocks by recrystallization is generally accompanied by a change in the chemical composition, that is, by loss and gain of chemical matter resulting in the so-called allochemical or metasomatic metamorphism, or simply metasomatism. One can often find statements to the effect that metamorphism without chemical change is the normal thing, and supposed cases of such isochemical [illegible] normal meta- [illegible] they are usually different, indicating that a relocation of the chemical constituents has taken place within the rock without affecting the total chemical composition. Therefore, in isochemical metamorphism the chemical elements have migrated over short distances and in allochemical metamorphism over large distances. In harmony with this nomenclature it has become customary, and is perhaps expedient in an attempt to survey and classify the various petrographic types, to distinguish between the normal metamorphism, with only small changes in the total chemical composition, and the metasomatic, or allochemical metamorphism, with large changes in the chemical composition. From a genetic point of view, however, this distinction is illogical and obscures the broader relations. Metamorphic rock always has undergone a change in

composition, for transportation of chemical matter and transfer of heat are not only concomitant with but essential parts of the earth processes constituting metamorphism. In geologic discussions the fact that sediments at the very early stages of metamorphism lose volatile constituents (water and carbon dioxide) and consequently regularly change their chemical composition has often been neglected. However, these changes should not be neglected, for the water and carbon dioxide in rocks are just as important as the many other rock-making constituents and serve to characterize the rocks in which they occur.

Large-scale chemical transport in and through large volumes of rock is the normal; gneiss-granites, migmatites, and mica schists all exhibit a chemical composition different from that of the preexisting rock complex. In a similar manner, the mineralogical composition changes with the changing chemical composition and with variations in temperature and pressure. Thermodynamic equilibrium is approached during the course of the metamorphism.

The mechanisms that produced the changes are not readily explained, but investigations bear out the conclusion that large-scale chemical changes in solid rocks did take place and that no melting occurred. This fact is further substantiated by studies which have been carried out on fossils in metamorphic rocks. So selective were the metamorphic processes that in many specimens the most delicate structures of the fossil organisms were preserved.

It is a reasonable position to base the discussion of recrystallization and evolution of metamorphic rocks on the postulate that free interchange of matter (atoms or ions) can take place over large volumes of rocks existing as essentially solid masses at temperatures well below 650°C. Above 650°C. magmas may form, but below this temperature rock magmas are rare or nonexisting. Thus there is a natural, though gradational, boundary between metamorphism and magmatism. *See* MAGMA.

Theory of metamorphism. The simplest kind of transformation in rocks involves changes occurring in individual minerals, that is, the polymorphic transformation of one mineral to another. Thermodynamically, changes occur in a mineral which cause a decrease in the free energy. Thus if a chemical compound such as silicon dioxide, SiO_2, occurs in a number of polymorphic forms such as quartz, tridymite, and cristobalite, then all polymorphic forms tend to transform into the one stable form which at a given temperature and pressure is characterized by the minimum free energy, for example, into tridymite at 900°C and 1 atm. This statement may be clarified by a graph (see illustration) and the useful, elementary thermodynamic relation,

$$A = U - TS$$

where A denotes the (Helmholtz) free energy of the mineral, U is the internal energy, and the quantity TS is the bound energy (T is the absolute temperature and S is the entropy). Thus in the equation, free energy equals internal energy less bound energy. *See* THERMODYNAMIC PRINCIPLES.

At absolute zero, the TS term vanishes, and the free energy becomes equal to the internal energy U of the crystal. Thus, at very low temperatures, the internal energy term is dominant and the polymorphic form with the least internal energy tends to be the stable one. With increasing temperature the TS term becomes increasingly important. It may happen, therefore, that because of the possibility of larger entropy in a second form, its TS term so reduces its free energy that, in spite of greater U, the differences between these two terms are greater than for the first polymorphic form. If this occurs the second form becomes the stable one, and the first form tends to transform into it. The temperature at which the free energies become equal is the transition temperature.

Analogously, the fact that a mineral assemblage changes into another shows that the new assemblage, as such, has a lower total free energy than the old. Under the conditions of the metamorphic environment, a new set of minerals, V, W, X, . . . , has a lower net free energy than the old set of minerals A, B, C, ... It is important to note that the energy drop represents the physicochemical force which drives the metamorphic process and that its potential comes from a difference in the free energy of the old set of minerals and the free energy of the new set of minerals.

The energy relations, therefore, are the controlling factors in metamorphism, and the geological agents determining these relations are mainly temperature and pressure.

It is worthy of note that, in general, the free energy of polymorphic transitions or of reactions between silicates is small, of the order of a few hundred or a few thousand calories per mole. This fact leads to formation or persistence of metastable

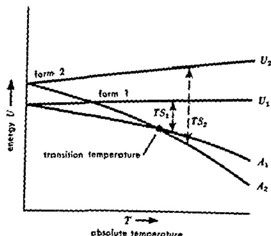


Diagram illustrating the relation $A = U - TS$ and showing how a high-temperature form (form 2), having high internal energy U_2 because of its high entropy S_2 at elevated temperatures, corresponds to the state of lowest free energy, A_2. (After M. J. Buerger, Am , 1948)

assemblages. Crystalline schists from various localities are known in which, for example, both sillimanite and kyanite, two forms of Al_2SiO_5, have been found together.

Types of metamorphism. Different kinds of metamorphism may be defined according to genetic criteria such as the geologic processes that were assumed to have caused the metamorphism, or the physical and chemical conditions that appear to have been predominant in determining the course of metamorphism. Using these criteria three general kinds of metamorphism are noted.

1. Dislocation, mechanical, or dynamic metamorphism is the result of pressure (or stress) along dislocations in the earth's crust. The deformed rocks commonly show marked zones of extremely fine-grained rocks, such as mylonites, whose structures are determined by crushing and movement of the grains without important recrystallization of old, or growth of new, minerals. This type of metamorphism is local and restricted in occurrence.

2. Contact or thermal metamorphism occurs in response to increased temperature induced by adjacent intrusions of magma. Chemical reconstitution of the rocks is due to magmatic exhalation; other conditions, such as confining pressure, exert subordinate influence.

3. Regional metamorphism, the most widespread type, is brought about by an increase in both temperature and pressure in orogenic regions, which are vast segments of the crust represented by the folded mountain ranges. The heat is produced partly by the downwarping and deep burial and partly by regional magmatism. Pressure is increased by deep burial and by directed pressure and shearing stresses accompanying the orogenic movements. *See* OROGENY.

Other systems of nomenclature are employed in classifying metamorphism and these may depart from the ideal simple categories noted above.

Autometamorphism is a term applied to chemical adjustment in newly congealed igneous rocks, brought about by a decrease in temperature. The residual hydrothermal fluids are then able to react with the igneous minerals.

Diaphtoresis, or retrograde metamorphism, is a general term denoting *any* metamorphism that takes place as a result of decreasing temperatures.

Static metamorphism (load metamorphism or geothermal metamorphism) is the term used to describe changes in rocks supposedly brought about by a regular increase in temperature and hydrostatic pressure through deep burial. However, such conditions are hardly sufficient to induce notable metamorphic effects in silicate rocks; but oceanic salt deposits, famous for their great varieties of minerals, have suffered profound changes in this way. Their metamorphism has been carefully studied and represents one of the best known physicochemical processes in geology. It has been shown that the famous salt deposits at Stassfurt, Germany, recrystallized under a superincumbent layer several kilometers thick and at about 80°C. *See* EVAPORITE (SALINE).

A distinction must be made between metamorphism in the superstructure or *oberbau* of the orogenic mountain chains and the infrastructure or *unterbau* lying below the usual depth range of regional metamorphism. Migmatization takes place as a result of the physicochemical conditions of the infrastructure. Thus migmatites are usually grouped with metamorphic rocks. *See* MIGMATITE. [T.F.W.B.]

Bibliography: See METAMORPHIC ROCKS.

Metaplasia

The appearance of a cell type at a site where it is not normally found. Metaplasia is usually attributed to either abnormal differentiation or maturation of the less mature cells normally present in this region. In a few instances, however, it has been suggested that adult cells may be transformed from one cell type to another.

Metaplasia from cuboidal or columnar (tall, prismatic in shape) cell types to squamous (flat, platelike) epithelium is commonly encountered in the bronchi as the result of chronic irritation and in the glands of the bladder as the result of vitamin A deficiency or chronic infection. Metaplasia of epithelium is also seen under conditions of extremely rapid growth, and is not uncommonly found in certain tumors. *See* EPITHELIUM.

Metaplasia of mesenchymal tissues is frequently encountered in association with necrosis, or cell death, injury with associated hemorrhage, and in scar tissue. The first stage is usually the formation of foci of calcification, and this is followed by the deposition of cartilage or bone. Metaplasia to osseous or cartilaginous tissues is also found in tumors. *See* DEATH. [W.F.F.]

Metasomatism

A process by which the chemical composition of rocks is altered. As originally defined by C. F Naumann, the term was applied to certain kinds of pseudomorphs, namely, those formed by chemical replacement at the expense of some original mineral. Every transformation of minerals by chemical replacement may now be designated as metasomatism, and in this sense, most of the newly formed constituents of the metamorphic rocks would be metasomatic minerals. *See* METAMORPHIC ROCKS; METAMORPHISM.

Metasomatism is fundamentally a change, atom for atom, in a rock mineral. It postulates a reaction between a solid mineral and a disperse or fluid phase. It corresponds to what several United States authors have called replacement. It may be well, therefore, to distinguish between mineral metasomatism and rock metasomatism. The mechanism of the mineral metasomatism in a metamorphic rock is the same as that of the rock metasomatism of large bodies; but in mineral metasomatism the fluid phase, which is generated of and by the metamorphic rock, would cause only a relocation of the chemical elements within the rock without changing the chemical composition of the whole rock. In the metasomatic rock, however, the fluid is intro-

duced into the rock and acts metasomatically, thereby changing the initial chemical composition of the rock. Thus the mineral metasomatism taking place in a metamorphic rock may be considered as a metabolism of rock, that is, a relocation of the chemical elements which have to migrate over but short distances; whereas true metasomatism of rocks should be restricted to comprise only those replacements by which the chemical composition of the whole rock, as well as that of the individual minerals, is essentially changed through an immigration of foreign chemical elements and an expulsion of other elements over long distances, up to several kilometers. Metasomatism in this sense would include, for example, the alteration of limestone into dolomite. Distinction may also be made between mutative, additive, and expulsive metasomatism.

Metasomatic processes take place over a wide range of temperature and pressure, and in rocks that are to all intents and purposes solid. They may be accompanied by a change in volume, but in most cases the replacements are effected without any appreciable change in volume (the "volume law" of Lindgren), and that is the reason why original structures and textures often are well preserved. In general, metasomatic reactions obey the law of mass action, and in order to effect replacement the introduced fluids must attain a minimum concentration. This explains why gradational passage from one type of rock into another is rare; sharp contacts are usual in metasomatic rocks.

A fine example of simple metasomatism is the formation of dolomite from limestone (calcite) in contact with sea water. Magnesium ions dissolved in the sea are slowly extracted through metasomatic reaction with limestone according to the schematic equation



$$2CaCO_3 + Mg^{++} \rightarrow CaMg(CO_3)_2 + Ca^{++}$$

Many dolomite rocks appear to have this mode of formation. The dolomitization of many coral reefs can be followed step by step (borings in the atoll Funafuti, Ellice Islands, Pacific). Phosphatization of limestones is another example.

Metasomatism is widespread in silicate rocks. In contact zones of igneous intrusions the replacements may affect not only the intruded rocks but also the crystallized magma from which the fluids are derived. Such magmatic fluids are often gaseous, giving rise to the so-called pneumatolytic metasomatism with formation of skarn rocks. These are lime-silicate felses (such as andradite garnet or hedenbergite rocks) which often contain special minerals rich in boron, fluorine, and sulfur. Valuable ore deposits are often associated with rocks of this character (copper, lead, silver). *See* ORE AND MINERAL DEPOSITS; PNEUMATOLYSIS; SKARN.

During the regional metamorphism metasomatism is ubiquitous. In all rocks the chemical changes are usually characterized by increase in the alkali metals, and decrease in magnesium, iron, and aluminum. *See* METAMORPHISM. [T.F.W.B.]

Metastable state

In quantum mechanics, an excited stationary energy state whose lifetime is unusually long (*see* EXCITED STATE; STATIONARY STATE).

The stationary states of atoms and nuclei, or indeed of any material system, ordinarily are computed ignoring the interactions between matter and the electromagnetic radiation field. When these interactions are included, a system in an excited state always has some probability of emitting radiation in the form of one or more photons, and thereby decaying to a lower or less excited state. For a metastable state, the probability of making such a transition is unusually small. Most atomic excited states decay with the emission of a single photon, in a time of the order of 10^{-8} sec. However, the necessity for angular momentum and parity conservation, forces the second excited state ($2S_{1/2}$) of atomic hydrogen to decay by simultaneous emission of two photons; consequently, the lifetime is increased to an estimated value of 0.15 sec. Similarly emission of a γ-ray photon by an excited nucleus usually occurs in 10^{-13} sec or less; however, the lifetime of one excited state of the In^{113} nucleus is about 100 min. Because the transition probabilities decrease rapidly with decreasing frequency, a low-lying excited state may have a lifetime longer than usual, and yet not be truly metastable; that is, emission from the state may not be hindered by any general requirement or selection rule, such as is invoked in the case of hydrogen. *See* ISOMERISM, NUCLEAR; QUANTUM THEORY, NONRELATIVISTIC. [E.C.]

Metatheria

An infraclass of therian mammals including a single order, the Marsupialia. The Metatheria are distinguished from the Eutheria (the placental mammals) by numerous characters. The full metatherian dentition is

$$i\frac{5}{4} \quad c\frac{1}{1} \quad p\frac{3}{3} \quad m\frac{4}{4} = 50$$

The braincase is small, the angular process of the [illegible]

theria arose from pantotheres in the Cretaceous at [illegible] were isolated by water barriers. *See* DENTITION; EUTHERIA; MAMMALIA; THERIA.

Metazoa

The *multice*[illegible] *animals that* [illegible] all of the animal [illegible] cept the [illegible] acellular [illegible] making up the [illegible] zoan body are [illegible] as [illegible] sues or organ [illegible]

assemblages. Crystalline schists from various localities are known in which, for example, both sillimanite and kyanite, two forms of Al_2SiO_5, have been found together.

Types of metamorphism. Different kinds of metamorphism may be defined according to genetic criteria such as the geologic processes that were assumed to have caused the metamorphism, or the physical and chemical conditions that appear to have been predominant in determining the course of metamorphism. Using these criteria three general kinds of metamorphism are noted.

1. Dislocation, mechanical, or dynamic metamorphism is the result of pressure (or stress) along dislocations in the earth's crust. The deformed rocks commonly show marked zones of extremely fine-grained rocks, such as mylonites, whose structures are determined by crushing and movement of the grains without important recrystallization of old, or growth of new, minerals. This type of metamorphism is local and restricted in occurrence.

2. Contact or thermal metamorphism occurs in response to increased temperature induced by adjacent intrusions of magma. Chemical reconstitution of the rocks is due to magmatic exhalation; other conditions, such as confining pressure, exert subordinate influence.

3. Regional metamorphism, the most widespread type, is brought about by an increase in both temperature and pressure in orogenic regions, which are vast segments of the crust represented by the folded mountain ranges. The heat is produced partly by the downwarping and deep burial and partly by regional magmatism. Pressure is increased by deep burial and by directed pressure and shearing stresses accompanying the orogenic movements. *See* OROGENY

Other systems of nomenclature are employed in classifying metamorphism and these may depart from the ideal simple categories noted above.

Autometamorphism is a term applied to chemical adjustment in newly congealed igneous rocks, brought about by a decrease in temperature. The residual hydrothermal fluids are then able to react with the igneous minerals.

Diaphtoresis, or retrograde metamorphism, is a general term denoting any metamorphism that takes place as a result of decreasing temperatures.

Static metamorphism (load metamorphism or geothermal metamorphism) is the term used to describe changes in rocks supposedly brought about by a regular increase in temperature and hydrostatic pressure through deep burial. However, such conditions are hardly sufficient to induce notable metamorphic effects in silicate rocks; but oceanic salt deposits, famous for their great varie- [illegible] ...ology). It has been shown that the famous salt deposits at Stassfurt, Germany, recrystallized under a superincumbent layer several kilometers thick and at about 80°C. *See* EVAPORITE (SALINE).

A distinction must be made between metamorphism in the superstructure or *oberbau* of the orogenic mountain chains and the infrastructure or *unterbau* lying below the usual depth range of regional metamorphism. Migmatization takes place as a result of the physicochemical conditions of the infrastructure. Thus migmatites are usually grouped with metamorphic rocks. *See* MIGMATITE. [T.F.W.B.]

Bibliography: *See* METAMORPHIC ROCKS.

Metaplasia

The appearance of a cell type at a site where it is not normally found. Metaplasia is usually attributed to either abnormal differentiation or maturation of the less mature cells normally present in this region. In a few instances, however, it has been suggested that adult cells may be transformed from one cell type to another.

Metaplasia from cuboidal or columnar (tall, prismatic in shape) cell types to squamous (flat, platelike) epithelium is commonly encountered in the bronchi as the result of chronic irritation and in the glands of the bladder as the result of vitamin A deficiency or chronic infection. Metaplasia of epithelium is also seen under conditions of extremely rapid growth, and is not uncommonly found in certain tumors. *See* EPITHELIUM.

Metaplasia of mesenchymal tissues is frequently encountered in association with necrosis, or cell death, injury with associated hemorrhage, and in scar tissue. The first stage is usually the formation of foci of calcification, and this is followed by the deposition of cartilage or bone. Metaplasia to osseous or cartilaginous tissues is also found in tumors. *See* DEATH. [W.F.P.]

Metasomatism

A process by which the chemical composition of rocks is altered. As originally defined by C. F. Naumann, the term was applied to certain kinds of pseudomorphs, namely, those formed by chemical replacement at the expense of some original mineral. Every transformation of minerals by chemical replacement may now be designated as metasomatism, and in this sense, most of the newly formed constituents of the metamorphic rocks would be metasomatic minerals. *See* METAMORPHIC ROCKS; METAMORPHISM.

Metasomatism is fundamentally a change, atom for atom, in a rock mineral. It postulates a reaction between a solid mineral and a disperse or fluid phase. It corresponds to what several United States authors have called replacement. It may be well, therefore, to distinguish between mineral metasomatism and rock metasomatism. The mechanism of the mineral metasomatism in a metamorphic rock is the same as that of the rock metasomatism of large bodies; but in mineral metasomatism the fluid phase, which is generated of and by the metamorphic rock, would cause only a relocation of the chemical elements within the rock without changing the chemical composition of the whole rock. In the metasomatic rock, however, the fluid is intro-

called trails. These phenomena may be observed by visual, photographic, telescopic, or radar techniques.

Meteor trajectories and radiants. When still some distance from Earth, meteors travel in elliptical or parabolic orbits about the Sun (*see* CELESTIAL MECHANICS). As they approach Earth their orbits are perturbed by Earth's gravitational field until the orbit is best described as a hyperbola about Earth's center. However, the arc of the hyperbola for the few tens of kilometers of trail may almost always be described as a straight line. Only in cases of extremely long and slow meteors is curvature of the trail detectable by the most accurate photographic and measuring procedures.

This straight-line trajectory is a great circle as seen projected on the celestial sphere. If the same meteor is observed from several different locations, the extension, backward along the trail, of the great circle observed at each station intersects the same point on the celestial sphere. This is the meteor's radiant point and is determined solely by the direction of motion of the meteoroid. In the same way, if a number of meteors traveling parallel to one another are observed from a single location, the backward extensions of their trails intersect at the common radiant point. Meteors with parallel motion and identical velocities must have identical orbits and, very likely, identical origins. The groups of meteors displaying these characteristics are called meteor showers (see illustration).

Meteor showers and comets. On the average, a visual observer may expect to see about five meteors per hour. At specific times of the year, however, this rate may increase greatly when Earth's orbit intersects the orbit of a stream of meteors Each year on August 12 in the early morning an observer in the Northern Hemisphere can see as many as 20 meteors per hour. Most of these have a common radiant point in the constellation of Perseus at right ascension (RA) 46° and declination (Dec) +58°. These meteors, named the Perseids after the constellation from which they appear to emanate, have nearly identical velocities of 60 km/sec before they are retarded by the atmosphere.

Again on December 13, at a few hours past midnight, a comparable number of meteors may be seen with a radiant at RA 113° and Dec +32° in Gemini. These are Geminids; their velocity is 36.5 km/sec.

The times and dates listed are for the maximum rate. The showers may be seen at a lesser rate on several days preceding and following the date of maximum and at other hours of the night when the radiant is above the horizon. The width of the stream, in the direction of Earth's orbit, must be several tens of millions of miles for the meteors to occur over several days. Also, because the showers repeat regularly each year and because the orbital period of the meteors is neither exactly 1 year nor even commensurable with Earth's period, the meteoric material must be distributed rather uniformly all about the stream's orbit.

There are sometimes condensations of meteoric material within an orbit. This was amply demonstrated in 1833 and 1866 when it was reported that the Leonids, ordinarily a shower of moderate activity, fell as thickly as snowflakes. Estimations of the maximum hourly rate in 1833 were 35,000. The average separation between particles in the stream at this time was about 20 km. The display of the Giacobinids on October 10, 1946 yielded visual counts of 1000 per hour at maximum activity. In most years this shower is nonexistent.

In 1866, G. V. Schiaparelli showed that the orbit of the Perseids corresponded with that of the Comet 1862III. Shortly after that, three other comet-meteor stream associations were made and the hypothesis that meteoric material represents debris from existing comets was firmly established. At present one can give 10 cases where an association assuredly exists (see the table). Nearly half the known meteor streams are included. The presumption is strong that the remaining showers also arose from comets which have either completely disintegrated or have been perturbed by planets into orbits too distant from the sun for observation.

Meteoric material is expelled from the comet along with gaseous material (*see* COMET). The velocity of departure from the comet nucleus is small enough to maintain a near identity of comet and meteoroid orbits, but large enough to assure that, in time, individual particles will be distributed rather completely about the comet's orbit. Effects of solar radiation are also important in perturbing the smallest particles into different orbits (*see* ZODIACAL LIGHT).

Occasional large observed rates for Leonids and Giacobinids represent examples of collisions with meteoroid streams that have not had sufficient time to disperse. The Giacobinids, debris from the Giacobini-Zinner Comet, are important only if Earth crosses the comet's orbit at about the time the comet is near that position of its orbit. The 1946 shower occurred 15 days after the comet made its closest approach to Earth's orbit. In 1939, although the approach to Earth's orbit was closer, the distance from comet to Earth was considerably greater (136 days late) and no shower was observed.

Observations. *Meteors are observed to determine* their brightness, velocities, and trajectories in the atmosphere and, through these latter two, their orbits about the Sun. To determine the trajectory, one must introduce a distance in the measures, either with a range measurement by radar or by triangulation between two stations observing the same meteor.

Visual observations. The apparent angular velocity of meteors ranges upward to tens of degrees per second. The eye is an exceedingly effective mechanism for detecting fast-moving objects; thus visual observations of meteors have been made extensively notwithstanding their low accuracy. A skilled observer can plot the observed great circle moti[illegible] with an error of about 1°. Two

Meteor shower and cometary associations

Shower	Date of max. Universal Time		Extreme limits		Radiant, ° RA	Radiant, ° Dec	Velocity, km/sec	Strength*	Parent comet
Quadrantids	January	3	January	1 4	230	+18	43	M	
Coma Berenicieds	January	17	January	13 23	186	+20	65	W	1913I
Virginids	March	13	March	5 21	183	+1	31	W	
Lyrids	April	21	April	20 23	270	+33	48	M	1861I
η-Aquarids	May	4	May	2 6	336	+0	64	M	Halley
Daytime Arietids	June	8	May June	29 18	44	+23	39	S	
Daytime ζ-Perseids	June	9	June	1 16	62	+23	29	S	
Daytime β-Taurids	June	30	June	24 6	86	+19	32	S	Encke
δ-Aquarids	July	30	July August	14 19	339	−11	43	M	
ι-Aquarids			July August	16 25	331	−5	31	W	
α-Capricornids	August	1	July August	17 21	309	−10	26	W	1948n
Perseids	August	12	July August	29 17	46	+58	60	S	1862III
Draconids (Giacobinids)	October	10	October	10	264	+54	23	Periodic	Giacobini-Zinner
Orionids	October	22	October	18 26	94	+16	66	S	Halley
Leo Minorids	October	24	October	22 24	162	+37	65	W	1739
Taurids	November	1	September December	15 15	51	+17	30	M	Encke
μ Pegasids	November	11	November	11	335	+21	16	W	1819IV
Leonids	November	17	November	14 20	152	+22	72	M	1866I
Geminids	December	14	December	7 15	113	+32	36	S	

* Shower strength (W, weak; M, medium; S, strong) is a relative measure of the number of meteors that may be observed visually, except for those daylight showers which are observed by radar.

observers at different stations can often determine the radiant with an error of only a few degrees, a sufficient accuracy for some types of analysis. However, the not surprising inabilities to estimate with accuracy the total angular path of the meteor and the time to traverse this path limit the value of visual observations in the determination of velocities.

Perhaps the most impressive results of visual observations are those by completely unskilled observers who, by chance, sight extremely bright meteors (commonly called fireballs). Many such observations of one fireball are often reported and enable an investigator familiar with the common errors of observations to estimate a space trajectory with sufficient accuracy to initiate a successful search for an unconsumed portion, the meteorite, that reached the ground.

Only with the best photographic techniques is it possible to record meteors fainter than can be seen by the naked eye. A few such photographs exist. Some telescopic data on fainter meteors have been obtained by a few excellent observers. However, the small field of view of a telescope limits the number of observations, and the high apparent speed with which the meteor traverses the field reduces the accuracy of the positional measures.

Photographic observations. Because neither time of appearance nor location of a meteor can be predicted, meteor photography is carried out on a patrol basis; that is, cameras are continuously open, and success is achieved when a sufficiently bright meteor chances to appear in the field. Usually two similar cameras, separated by a baseline of 20-80 km, are operated simultaneously. These are directed at the same region in the atmosphere at a height of about 90 km. Both cameras photograph the same meteor and the two photographs supply the necessary trajectory data. To determine velocities, a time scale is introduced in the meteor trail

The most common device is a rotating shutter which alternately exposes and occults the objective or the focal plane at known intervals. Interruption rates from 20 to 60 per second are desirable.

Cameras suitable for meteor photography possess (1) a wide field of view to patrol a large region of the sky, (2) a large aperture to gather as much light as possible at any instant during the meteor's flight, and (3) a short focal length to decrease the angular rate of the image and, consequently, to concentrate the light on a small area of the photographic emulsion.

The most successful instrument in these respects is the Baker Super-Schmidt meteor camera. Its focal length is 8 in., and its effective aperture is approximately 10 in. With fast blue-sensitive emulsions, slow meteors of 5th magnitude and fast meteors of 3rd magnitude can be photographed. The field of view of the camera is 57°. Approximately four meteors are photographed during each hour of exposure.

Radar observations. Radar observations of meteors depend on the reflection, or backscatter, of the radar pulse from the ionized column of gases formed as the meteor passes through the upper atmosphere. The echo reaches a pronounced maximum when the meteor trail is at right angles to the radar beam. The distance to this point is determined by the time lapse between the transmitted and received pulses. If a continuous-wave transmission is used, the pattern of the reflected signal shows oscillations as a result of the diffraction produced by the growing line source. A measurement of the rate of oscillation, together with a measurement of radar range, is sufficient to determine the velocity.

Radiants of meteor showers are determined by a series of observations with the radar directed towards the horizon. When the detection rate reaches a maximum, the meteor trajectories are at right angles to the beam. The center of the beam then represents the pole of the great circle containing the radiant; the radiant is 90° from the beam. Observations at differing azimuths yield two great circles, the intersection of which must be the radiant.

Spectroscopic observations. The first photographic observation of a meteor spectrum was made in 1897; since then only a few hundred more have been obtained. Bright meteors are required; a single camera equipped with a transmission grating or prism may acquire less than 10 meteor spectra per year.

Spectra show that the luminosity arises primarily from emission lines of neutral atoms and once-ionized atoms. The predominant lines of ionized calcium and much of the light produced by iron atoms account for the blueness of all meteors. The color index of photographic meteors, on the average, is a −1.8 magnitude (*see* MAGNITUDE, STELLAR).

The only molecular bands detected in meteor spectra arise from the first and second positive groups of atmospheric nitrogen (*see* MOLECULAR STRUCTURE AND SPECTRA). A comparison of elements responsible for the spectra and the elements known to occur in meteorites proves that most photographic meteors are stone rather than iron. Only a few spectra exist that may be attributed to iron.

The highest observed excitation potentials (upper states) in meteor spectra are due to hydrogen and ionized silicon These are between 12 and 13 electron volts. The degree of excitation and ionization increases as the meteor progresses into the atmosphere. High-velocity meteors show greater excitation and ionization than those of low velocity Many bright meteors display sudden and short-lived increases in brightness caused by the detachment of small droplets or particles from the meteoroid In these bursts, the degree of excitation always decreases.

Masses, magnitudes, numbers of meteors. The brightness of a meteor is generally described in terms of its apparent stellar magnitude at maximum light. The number of visual meteors increases with decreasing brightness, although the observed rate of meteors near the visual limit is low because the sensitivity of the eye is not uniform over the field of vision. It is estimated that an experienced observer sees fewer than 10% of the meteors in his field between the 5th and 0th magnitude and sees essentially all meteors brighter than 0th magnitude

Visual and photographic observations show that the number of meteors between magnitudes M and M + 1 is about 3.5 times the number between magnitudes M − 1 and M. About 10^8 meteors of the 5th and fainter magnitudes strike Earth daily.

Although some uncertainty is still present in the mass determination of meteors, an order of magnitude estimate may be given. A 30-gram meteoroid with velocity of 30 km/sec will produce a 0th magnitude meteor.

The total mass of meteors, including those fainter than the visual limit, may be obtained if one cares to risk an extrapolation of the number-magnitude law to faint meteors. Such an estimate gives [illegible]

[illegible]

the wake, a short appendage behind the meteor proper. The wake appears on a photograph as luminosity which partially fills in the breaks of the trail introduced by the rotating shutter. Spectra of the wake show the material to be compatible with the meteor itself although with the lower excitation expected of a smaller particle.

The wake may be caused either by meteoric gases or by small fragments which have been retarded with respect to the parent body. Existing

observations are insufficient to assign the responsibility to one mechanism or the other. Quite probably both play a role.

The second special case of meteor luminosity is the meteor train. This is a relatively long-persisting luminosity which may be seen for some seconds, or more rarely, some minutes after the meteoroid has passed. It is a streak of light that marks the path of the meteor. The phenomenon is most frequently produced by high-velocity meteors.

Spectra of trains are exceedingly rare and most of our information is derived from spectrographic observations with a camera equipped with a rotating shutter. If the brightness of a given spectral line is nearly constant in both the exposed portion of the meteor trail and the portion occulted by the shutter, it must result from luminosity that persisted through many shutter rotations.

P. M. Millman first obtained examples of such persistent luminosity, the lines were attributed to *magnesium, calcium,* iron, and, probably, sodium. More recently a line at 5577 A was photographed in a number of train spectra. This radiation arises from a metastable state of the oxygen atom and is also *responsible* for green aurorae (*see* AURORA). The mechanism of train production, other than by oxygen, is not yet understood. The half-life of the auroral oxygen line is about 1 sec and this *may* solve *the problem* for the short-duration trains. There is some evidence that this type of train is produced more readily when the 5577-A radiation from the sky as a whole is pronounced.

Trains persisting *for* some seconds or more frequently depart *from linearity* as the upper atmospheric winds blow them about. Time-lapse photography of trains is employed to *determine* the winds at altitudes of between 80 and 100 *km*.

Nature of meteoric matter. Information on the composition and *structure* of meteoric material is derived *directly from* the analysis *of meteorites* or indirectly from observations *of meteor* phenomena in the atmosphere. The first *source,* in which the investigator may actually *handle* the body he studies, is unique in the field of astronomy, and would appear to reduce the problem *to* that of finding a sufficient number *of* meteorites for statistical study. However, in addition to *stony* and iron meteorites, at least one *other class* of meteoric material exists. It is represented in *essentially* all of the ordinary visual meteors.

Apart from the elements *they contain,* visual meteors may bear no *similarity* to meteorites. There is good evidence that *most meteoroids fragment* as they pass through the atmosphere *and* some evidence that nearly complete *disruption* of certain meteoroids occurs whenever the dynamic pressures on the meteoroids exceed 0.5 *psi.* The general fragmentation may come *about* by molten *droplets* blowing off a substantial meteoroid, *but* the destruction of an entire meteoroid by so small a pressure, and not by thermal *effects,* is indicative of extreme fragility.

Other observations indicate that meteoroids may have an over-all density much less than that of the stone or iron meteorites (meteoroids as they finally reach the ground). The density of meteoroids may possibly be less than that of water. It is presumed that the low density and extreme fragility are due to porosity of the meteoric material.

On the basis of F. L. Whipple's icy-conglomerate model of comets one would expect porosity which results from the evaporation of the volatile ices from between the solid structure.

Artificial meteors. A new approach to meteor physics was initiated on October 16, 1957, when three artificial meteors were flown to an altitude of 80 km in a rocket over New Mexico. The meteors were formed by high-velocity ejection of the liners of shaped charges, an explosive device designed *to direct* the energy of the explosion along a specific line (*see* SHAPED CHARGE).

Two of the charges were successfully fired, and one yielded some information. The fraction of a gram of micron-sized particles of aluminum ejected by this charge reached a velocity of 15 km/sec and was clearly recorded on photographic films as a trail *about* 1000 *meters* long. By the end of their visible trail the particles were essentially completely vaporized and the remains brought to a stop by atmospheric drag. Iron liners producing larger *particles are* required to simulate the natural meteor sufficiently well to determine the luminosity coefficient of the meteoric process. The high velocity required for vaporization and luminosity may also be obtained from rockets, missiles, and satellites reentering Earth's atmosphere. [R.E.Mc.]

Bibliography: R. L. F. Boyd and M. J. Seaton (*eds.*), H. S. W. Massey, consultant, *Rocket Exploration of the Upper Atmosphere,* 1954; T. R. Kaiser (ed.), *Meteors,* 1955; A. C. B. Lovell, *Meteor Astronomy,* 1954; S. K. Mitra, *The Upper Atmosphere,* 2d ed., 1952; J. G. Porter, *Comets and Meteor Streams,* Int. Astrophysics *Ser.,* vol. 2, 1952; F. G. Watson, *Between the Planets,* rev. ed., 1956.

Meteorite

A solid body of subplanetary size, either in space or fallen to the earth, which retains some of its primordial characteristics. Although *they* are rare on this earth, meteorites make up a large proportion of the solid matter in the *universe.* Usually meteorites are named after the town or post office nearest to which they fell or were *discovered.* For a discussion of meteoroids, their trajectories and velocities, and techniques of observation, *see* METEOR.

Fall of a meteorite. Falling meteorites are sometimes visible over thousands of square *miles* and are heard over hundreds of square miles. These spectacular events have startled man throughout the ages.

The fact that solid matter falls from the sky was not established until 1768, when documentary evidence of the fall of the Luce meteorite was pre-

sented to the Royal Academy of Science in Paris. The first witnessed meteorite to fall and be preserved in the United States fell at Weston, Connecticut, on December 14, 1807.

Meteorites enter the atmosphere so fast that in their collision with the air they develop a high temperature. The leading face of the meteorite and the air in front of it becomes hot. Most of the material that is evaporated, melted, or mechanically lost from the meteorite comes from its forward face. Thus a meteorite changes in size and shape as it falls.

Recent studies on the distribution of helium-3 in a 1000-lb iron found near Grants, New Mexico, showed that it was angular before it entered the earth's atmosphere; and because helium-3 decreases with depth from the surface, and the zones of equal helium-3 content are concentric with its preatmospheric shape, this iron lost more than 50% of its mass in passing through the atmosphere

The particles removed from a falling meteorite are swept aft into the turbulent air and form the sustained trail which makes a conspicuous luminous streak in the night sky. When a meteorite breaks in the air, the pieces scatter, but because the trajectory of the meteorite is inclined towards the earth, the fragments fall within an elliptical area Larger masses retain more of their velocity and thus are carried to the far end of the fallout area, while the smaller pieces are slowed by the air and fall at the near end of the strewn field.

Since 1800, falls of about 627 stony, 29 iron, and 10 intermediate types of meteorites have been witnessed. Also, 9 doubtful falls are reported, including 8 iron meteorites and 1 stone one. The maximum number of falls in a year was 13 in 1930, and in 6 of the years, no meteorite falls were reported. Four of these years were before 1817; the others were 1832 and 1880.

More meteorites are seen to fall than are found Some meteorites are essentially consumed in the air. Others fall in sparsely settled areas or in water The population distribution and the newspaper and communication facilities influence the recording of this information. Sometimes years pass before a fall is reported. Although statistics about meteorite falls are incomplete, the records about the weights of the fallen materials are even less accurate. Meteorite falls are seldom completely recovered, and one shower may overlap a previous one.

Distribution. The known meteorites are not evenly distributed over the land masses of the earth Some of the reasons for this are unequal population distribution, the nature and use made of the land, and the interest people have in such events. About 1700 meteorites are known, but an indefinite number are paired falls (different names given to scattered pieces of the same shower). Table 1 lists meteorites reported from several selected areas to illustrate the unequal distribution.

About 33% of the world's meteorites are reported from the United States, which has an area of less than 6% of the land area of the world.

Table 1. Known meteorites per square mile for seven countries

Country	Area, mi²	Total meteorites	Area per fall, mi²
Australia	2,973,805	103	28,875
British Isles	121,351	21	5,778
France	212,766	58	3,668
India	1,269,312	123	10,319
Japan	142,754	39	3,660
Mexico	760,175	58	13,106
United States	3,615,223	620	5,831

Table 2. Stony, intermediate, and iron meteorites, and witnessed falls of iron meteorites reported in five countries

Country	Known meteorites			Witnessed falls of irons	Irons in country, % of total meteorites
	Stones	Intermediate	Irons		
India	117	2	4	4	3.2
Italy	19	1	1	1	4.7
Spain	21	1	4	2	15.3
Russia	92	3	27	4	22.1
United States	338	19	263	3	42.4

Table 2 records the types of meteorites reported from five selected countries, and the witnessed falls of iron meteorites in these countries. In India and Italy, all the irons are witnessed falls; in Spain only 50% were witnessed. In Russia and the United States, about 15 and 1%, respectively, were seen to fall The abundance of irons in North America is probably due to the fact that the aborigines on this continent made little use of iron; in areas where old cultures existed, the iron meteorites were used by those peoples.

Geologically old falls. The statistics previously given are primarily for the meteorite falls of recent geologic times; however, a few of these may have fallen in earlier ages. The recent announcement of the wide distribution of nickel-iron spherules in the Pacific deep-sea sediments shows that numerous meteorites passed through the atmosphere in Tertiary times This discovery strengthens the evidence that some of the meteorites recovered from certain beds actually fell when these sediments were being formed.

The Aggie Creek, Alaska, iron was found at bedrock under 13 ft of alluvial material. The Keyes, Oklahoma, stone was said to have come from a Pleistocene formation, and the Klondike, Canada, iron came from beneath 65 ft of sediments of this same age. The Sardis, Georgia, iron probably fell in Miocene times when the beds from which it was dug were formed.

If the iron encountered in drilling 1525 ft into Eocene beds in Zapata County, Texas, is authenticated, it will be the oldest known fall.

Velocity and size of meteorites. The velocity of a meteorite coming directly toward the earth is obtained by adding to the apparent velocity of

meteorite 18.5 mi/sec, the speed at which the earth rotates around the sun. Such a colliding meteorite has the highest possible speed; therefore, a large one would hit the earth with great force. The slowest meteorites are those which overtake the earth or spiral in. To calculate their speed, 18.5 mi/sec is subtracted from the velocity of the meteorite. Meteorite velocities may vary from about 7.5 to 45 mi/sec. The duration of a meteorite's flight in the atmosphere depends on how it approaches the earth

The largest recovered meteorites are irons, which are less likely to break in flight than a stony one. The heaviest individual meteorite is the 60-ton iron from Hoba, South-West Africa. The largest stony meteorite weighs 2300 lb and fell on February 18, 1948, in Norton County, Kansas. The largest meteorite from the United States is the Willamette, Oregon, iron, which weighs about 13.5 tons. The smallest meteorites are dustlike particles.

Shapes of meteorites. Falling objects assume a more or less fixed position. Thus meteorites often have front and rear faces. The fusion crust on the front sometimes displays flow lines, but more often meteorites have depressions which are elongated in the direction of air flow. Generally, the leading face is dome-shaped and the rear side is flatter and irregular. Occasionally, an accumulation of fusion crust occurs immediately behind the front face (Figs. 1 and 2).



Fig. 1 Freda, North Dakota, meteorite showing opposite sides (a) and (b). Arrangement of ridges on this meteorite shows it fell in a fixed position.



Fig. 2. Amherst No. 2, Nebraska, meteorite. Leading face of a stony meteorite is frequently domed and has lines radiating from the center to the edge. The lines are deflected around the pits; therefore the depressions were formed before the lines; however, both the depressions and radial markings were made in flight. Similar depressions occur in many meteorites.

Table 3. Average composition of iron, pallasite, and stony meteorites, and terrestrial igneous rocks, weight %

Component	Meteorites			Terrestrial igneous rocks
	Iron	Pallasite	Stone	
Fe	90 85	55.33	15 5	5 00
Ni	8.5	5.43	1.10	0 008
Co	0 6	0 30	0 08	
Cu	0.02		0 01	0 007
P	0 17		0.10	0 118
S	0 04		1 82	0 052
C	0 03		0 16	0 032
O		18 55	41.0	46 60
Mg	0 03	12.33	14 3	2 09
Ca	0.02		1.8	3 63
Si	0 01	8 06	21.0	27 72
Na			0 8	2 83
K			0 07	2 59
Al			1 56	8 13
Mn	0.05		0 16	0 10
Cr	0 01		0 40	0 02
Ti			0 12	0 44

Meteorites have various shapes. Stony meteorites usually are rounded, irregular bodies, whereas the irons are more angular. Irons may have conical, rod, ring, or flat forms. Shortly after a meteorite falls, weathering processes begin to change their color, and eventually many change their shape. Corrosion is more active on the underside of an iron than on the exposed side. Meteorites containing iron chloride (lawrencite) are unstable. Thus the shapes of both stony and iron meteorites change if long exposed to weathering conditions.

Much attention has been given to the shapes of meteorites, but features such as flow lines, pits, and surface contour are equally important because these show how the mass reacted in the air during its high speed flight.

Temperatures of falling meteorites. Meteorites are relatively cold until they enter the atmosphere, but once in the atmosphere, the air in front of a falling meteorite and the forward face of the mass become hot. The air resistance which heats their surfaces also retards their velocity. Thus, small meteorites are retarded so much by the end of their flight that their surfaces are being cooled rather than heated.

Many old reports indicate that meteorites were hot when they fell, but such claims have not been verified. Meteorites have fallen on dried leaves without scorching them and have struck buildings without starting fires. The Dhurmsala, India, stone fell on July 14, 1860, and the Colby, Wisconsin, stone fell on July 4, 1917; both were recovered immediately and were cold; frost formed on the Colby stone. However, a large meteorite has enough mass that it can retain high enough velocity to strike the earth while its surface is being heated. Surfaces of

meteorites which produced large craters probably were hot when they fell.

Although meteorites are covered with a thin crust when they fall, their interiors are not heated when they enter the atmosphere or during their fall. The heat generated during the fall rarely penetrates the mass more than 1 cm. This is confirmed by laboratory studies on the transformations in metals and silicates. The material softens when heated and is easily removed by the air. The process by which the mass of a falling object is lost is called ablation.

Composition. The elements occurring in terrestrial rocks are also found in meteorites and combine to form similar minerals. However, there are fewer meteoritic than terrestrial minerals because many of the latter crystallize from solution and others are alteration products. Water occurs in meteorites, but few meteoritic minerals are hydrous.

Meteorites have more complex structures and the meteorite minerals are less homogeneously distributed than the minerals in terrestrial rocks. This makes it difficult to obtain an average composition of a meteorite. Since there are fewer analyses of meteorites than terrestrial rocks, the average composition of the types of meteorites is less accurately known than the average composition of the earth's crust (Tables 3 and 4).

Table 4. Meteorite minerals

Group and name	Formula	Crystal system
Native elements		
Kamacite[a]	NiFeCo	Isometric
Taenite[a]	NiFeCo	Isometric
Copper	Cu	Isometric
Diamond	C	Isometric
Graphite	C	Hexagonal
Carbides		
Cohenite	$(Fe,Ni,Co)_3C$	Isometric
Moissanite[b]	SiC	Hexagonal
Nitrides		
Osbornite	TiN	Isometric
Sulfides		
Troilite	FeS	Hexagonal
Oldhamite	CaS	Isometric
Daubreelite	$FeCr_2S_4$	Isometric
Phosphides		
Schreibersite	$(Fe,Ni,Co)_3P$	Tetragonal
Rhabdite	$(Fe,Ni,Co)_3P$	Tetragonal
Chlorides		
Lawrencite	$FeCl_2$	
Oxides		
Magnetite	$FeFe_2O_4$	Isometric
Chromite	$FeCr_2O_4$	Isometric
Quartz	SiO_2	Hexagonal
Tridymite	SiO_2	Orthorhombic
Pyroxenes		
Enstatite	$MgSiO_3$	Orthorhombic
Clinoenstatite	$MgSiO_3$	Monoclinic
Bronzite	$(Mg,Fe)SiO_3$	Orthorhombic
Clinobronzite	$(Mg,Fe)SiO_3$	Monoclinic
Hypersthene	$(Fe,Mg)SiO_3$	Orthorhombic
Clinohypersthene	$(Fe,Mg)SiO_3$	Monoclinic
Diopside	$CaMgSi_2O_6$	Monoclinic
Hedenbergite	$CaFeSi_2O_6$	Monoclinic
Augite	$CaNa(Mg,Fe,Al)(Al,Si)_2O_6$	Monoclinic
Olivine		
Olivine	$(Mg,Fe)_2SiO_4$	Orthorhombic
Forsterite	$(Mg)_2SiO_4$	Orthorhombic
Feldspars		
(Plagioclase feldspars) Albite-anorthite	$NaAlSi_3O_8$ / $CaAl_2Si_2O_8$	Triclinic / Triclinic
Maskelenite	Fused feldspars (?)	Amorphous (?)
Orthoclase	$KAlSi_3O_8$	Monoclinic
Hydroxysilicates		
Chlorite	$(Mg,Fe)_5(Al,Fe)_2Si_3O_{10}(OH)_8$	Monoclinic
Serpentine	$Mg_3Si_2O_5(OH)_4$	Monoclinic
Phosphates		
Apatite	$Ca_5(F,OH,Cl)(PO_4)_3$	Hexagonal
Merrillite	$Na_2Ca_3(PO_4)_2O$ (?)	Hexagonal

Table 4. Meteorite minerals (Cont.)

Group and name	Formula	Crystal system
Carbonates		Hexagonal
Breunnerite	$(Fe,Mg)CO_3$	Hexagonal
Calcite	$CaCO_3$	
Hydrocarbons, amorphous, carbonaceous, and organiclike compounds[c]		
Nitriles	$-C{\equiv}N$	Amorphous
Substituted imines	$>C{=}N{-}\overset{\mid}{\underset{\mid}{C}}-$	Amorphous
Amorphous carbon	$-\overset{\mid}{\underset{\mid}{C}}-\overset{\mid}{\underset{\mid}{C}}-$	Amorphous
Nitroso compounds	$R{-}NO^{d}$	Amorphous
Covalent sulfonates	$R{-}O{-}SO_2{-}R$	Amorphous
Aliphatic hydrocarbons	$RC{\equiv}CH$	Amorphous
Amines	$>N{-}H$	Amorphous
Water	$H{-}O{-}H$	Amorphous

[a] Kamacite and taenite are alloys of nickel and iron. See Figs. 5 and 6 for the occurrence and structure

[b] Moissonate has been reported but not verified.

[c] Chemical groups identified by infrared analysis; exact chemical formula unknown.

[d] R, hydrogen or organic groups.

Stony meteorites contain less oxygen, silicon, aluminum, calcium, sodium, and potassium, and much more iron, nickel, cobalt, sulfur, and magnesium than the average igneous rocks. Iron meteorites are chemically different from terrestrial iron and artificial products.

Structures and types of meteorites. Stony meteorites may be compared to volcanic rocks, yet there are characteristic differences. The minerals in both substances are chemically identical, but stony meteorites have unique properties which set them apart from terrestrial rocks. Stony meteorites are never bedded, nor do they contain conspicuous amounts of quartz. Frequently they contain small inclusions of nickeliferous iron.



Fig. 3. Tieschitz, Czechoslovakia, meteorite. Structure of stony meteorites is best seen when transmitted light passes through a thin section of the stone. This chondrite contains whole and fragmentary chondrules in a ground mass darkened by the abundance of glassy siliceous material and opaque metallic minerals, troilite (FeS), and metal. The large chondrule consists of three constituents, olivine (banded cheveron pattern), pyroxene (acicular), and some dark glassy material.

The two major types of stony meteorite are distinguished by the presence (chondrite) or absence (achondrite) of peculiar spherical inclusions called chondrules. These bodies are generally composed of pyroxene or olivine, and are sharply delineated from the surrounding matrix; however, they are not always easily freed from the matrix. When examined in cross section, thondrules often show a diverging fibrous structure from some eccentric point, a banded structure, or crystalline fragments of one or more minerals and occasionally fragments of chondrules. The minerals are often banded with a dark, fine-grained, glasslike material and separated by this material (Fig. 3). A classification of stony, iron, and intermediate-type meteorites (with principal constituent) follows:

I. Stony meteorites with or without small metallic inclusions
   - *A.* Chondrites (usually when Ni increases in metallic inclusions, ferrous iron increases in the magnesium silicates; to each group, distinctions such as gray, white, veined brecciated are applied)
     1. Enstatite chondrite
     2. Bronzite chondrite
     3. Hypersthene chondrite
   - *B.* Achondrites (grouped by calcium content)
     1. Calcium-poor achondrites
        - Aubrites (enstatite)
        - Ureilites (clinobronzite and olivine)
        - Amphoterites (bronzite and olivine)

Rodites (bronzite and olivine, brecciated)
Diogenites (hypersthene and olivine)
Chassignites (mostly olivine)

2. Calcium-rich achondrites
Angrites (mostly augite)
Nakhlites (diopside-olivine)
Eucrites (clinohypersthene and anorthite)
Sherghottites (clinohypersthene and maskelynite)
Howardites (hypersthene and anorthite)

II. Stony-iron meteorites (intermediate types)
Pallasites (olivine in iron)
Sorotiites (troilite in iron)
Siderophytes (bronzite, tridymite, and iron)
Lodranites (bronzite-olivine and iron)
Mesosiderites (hypersthene-anorthite and iron)

III. Iron meteorites (essentially nickel-iron masses with occasional silicate inclusion)
A. Hexahedrites and nickel-poor ataxites; nickel content less than 6% (Fig. 4)
B. Octahedrites; nickel content 6–14% subdivided according to the widths of the kamacite bands into coarsest, coarse, medium, fine, and finest octahedrites (nickel content increases in same order). Octahedrites, when polished and etched, display a Widmanstätten pattern (Fig. 5). The wide lamellae are kamacite, low nickel-iron phase (α-iron); taenite forms thin lamella at edge of kamacite bands, high nickel-iron phase (γ-iron); plessite, a mixture of kamacite and taenite sometimes occurs between kamacite bands
C. Nickel-rich ataxites; nickel content usually 14–30% (Fig. 6)

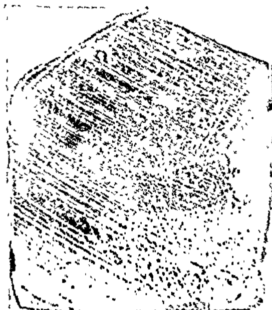


Fig 4 Edmonton, Canada, meteorite. Typical etched pattern of a hexahedrite. Uniform system of diagonal markings (Neumann lines) shows that this iron is a single crystal. Neumann lines disappear when the metal is heated to about 800°C, proving that the iron has not been hot since the lines formed. Lines which intersect at right angles are parallel to the direction of a cube; and kamacite, α-iron, cleaves in this direction. Octahedrites, hexahedrites, and many metallic inclusions in stony meteorites show Neumann lines; thus, it is presumed they originated before the meteorite entered our atmosphere.



Fig. 5. Grants, New Mexico, meteorite. The Widmanstatten pattern consists of an orderly arrangement of two phases of iron and nickel, phases where the kamacite bands parallel the octahedral directions in a cube. Taenite, the γ-iron phase, is invisible at the edge of the kamacite. Plessite, the dark angular areas between the kamacites, consists of unresolved kamacite and taenite. The round inclusion is troilite (FeS) and the long narrow ones are schreibersite, iron, nickel phosphide.

Origin of meteorites. Meteorites are physically and chemically complex bodies and no simple theory explains their origin. The largest class of stony meteorites, chondrites, consists of silicate minerals which after forming were transported to where they accumulated. During the consolidation of these fragments or afterwards changes occurred. The pieces became bonded together, metallic constituents were deposited around many of the chondrules, metamorphic changes occurred in the silicates, and secondary veins developed.

Most of the iron meteorites consist of single crystals. First the metal solidified from a melt, and as the temperature decreased, transformations took place. If the composition of the melt was above 6% nickel, two phases separated and a Widmanstätten structure formed.

The intermediate types of meteorite, such as pallasites, contain minerals similar to those found

stony meteorites and iron similar to the metal phase in stony and iron meteorites. Therefore one can postulate that all meteorites were formed by a continuous sequence of events.

The similarity between meteorites and various zones within the earth indicate that meteorites and the earth originated in a similar manner. From this and other astronomical evidence, it is thought that meteorites probably originated within our solar system. They may represent the debris from the collision of asteroids. No attempt is made in this explanation to outline the chemical processes that produced the minerals or the pressures and temperatures involved in the process.

Age of meteorites. Various methods of determining the age of meteorites have given similar results. The most accurate are the Pb^{207}/Pb^{204} and the Ar^{40}/K^{40} methods, which give ages of about 4.5×10^9 years. Apparently the age of the earth is the same as the age of meteorites. *See* EARTH (AGE OF).

Identification of meteorites. Meteorites have characteristics which set them apart from terrestrial rocks; however, these points will be useful in recognizing only the common meteorites. To recognize an unusual variety requires experience and special equipment; thus, the finder of a possible meteorite needs the service of a specialist. The U.S. National Museum in Washington, D.C., collects meteorites and examines specimens believed to be meteorites.

All meteorites are rare; thus, the finder should compare his find with the rocks of that immediate area. Iron meteorites are heavy, and because iron seldom occurs in nature, a heavy object should be investigated. Stony meteorites also are relatively heavy, but they usually have a crust over the surface which is different from the interior. If the crust is missing, the gray interior may have small metallic inclusions, small rounded silicate masses, chondrules, or both. Weathered meteorites generally are brown. Iron meteorites are strongly magnetic, and most stony meteorites contain enough iron to be attracted to a magnet. Surface features of meteorites are shown in Figs. 1 and 2.



Fig 3. [illegible], North Dakota, meteorite. Internal structure revealed when polished surface is etched; in this case, a characteristic ataxite structure is shown. The large inclusions, kamacite and $FeFe_3P$ eutectic, are enclosed in an acicular ground mass of kamacite and taenite. Light areas around kamacite inclusions represent concentrations of kamacite.

Table 5. Famous meteorite craters

Crater	Location	Approximate dimensions: Diameter, ft	Approximate dimensions: Depth, ft
Aouelloul	Western Sahara, Africa	817	
Arizona (Canyon Diablo)	Winslow, Coconino County, Arizona	4,100	600
Box Hole	Central Australia	575	
Chubb[a]	Ungava, Quebec, Canada	10,560	
Dalgaranga	Western Australia	225	15
Haviland[b]	Kiowa County, Kansas		
Henbury[c]	Central Australia	660	40
Kaalijarv	Estonia	327	16
Merewether[a]	Northern Labrador	455	125
Odessa	Ector County, Texas	530	18
Sikhote-Alin[c]	Maritime Province, Eastern Siberia, U.S.S.R.	75	20
Tunguska River[c]	Central Siberia, U.S.S.R.	163	18
Wabar	Southern Arabia	300	40
Wolf Creek	Western Australia	2,800	160
Campo del Cielo[d]	Northern Argentina		

[a] No meteorite fragments have been discovered.
[b] Small depressions.
[c] Group of craters. Dimensions given are those of the largest crater.
[d] Doubtful. Probably Indians dug these pits and transported the metal to this locality.

Collectors prefer meteorites which have not been broken, treated with reagents, heated, or damaged by careless handling. Such handling detracts from the importance of the specimen and makes it less valuable to science. Thus it is important to notify a museum when a possible meteorite is found.

Meteorites contain nothing which can be profitably extracted. Their only value is the scientific information they contain. Meteorites have no fixed values. To appraise a specimen properly, a number of factors should be considered: size, number of specimens available, type, presence of unusual features, and the degree of preservation.

Meteorite craters. Apparently most of the meteorites entering the atmosphere are small and are retarded by the air, so their terminal energy buries them only a few inches in the ground. Large meteorites are not slowed down, and therefore make craters. The oldest and best-known meteorite crater is that formed by the fall of the Canyon Diablo meteorite in Arizona (Fig. 7).

The most characteristic feature about such scars is the upturned rim. This was formed when the rocks rebounded after impact. The temperature de-

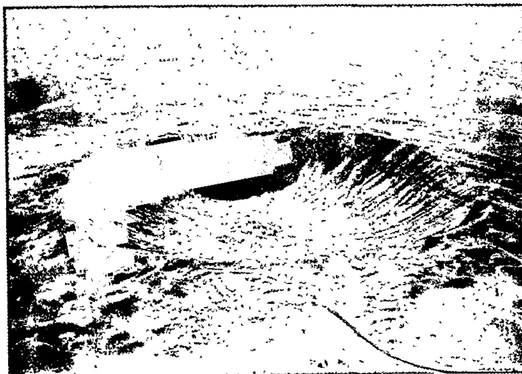


Fig. 7. Canyon Diablo, Arizona Crater. Facing south. (Photograph by John Farrell)

veloped vaporized most of the meteorite, and the shattered material was thrown out of the hole. Some of the well-known meteorite craters are listed in Table 5. [E.P.H.]

Meteorological instrumentation

Apparatus and equipment used to obtain quantitative information about the weather. This information includes the state of the atmosphere in such aspects as motion, energy exchange, gaseous composition, solid and liquid particle content, electrical activity, and others, as well as their composite effects in various combinations. Air temperature and pressure, which describe the state of the atmosphere, are obtained with simple instruments such as thermometers and barometers. However, visibility influenced by the amount and size distribution of liquids or solids in the atmosphere and the amount and direction of illumination is an example of a meteorological parameter which is not so easy to measure directly, but is expressed as the distance to which objects can be recognized. Any instrument able to measure useful properties of the atmosphere is potentially a meteorological instrument. Ordinarily, however, the expression is used to indicate instruments in use for routine weather observations.

Classes of observations. Four types of observation are made: climatological, synoptic, operational, and research.

Climatological observations. These measure the broad features of the weather at a given location over a long period of time.

Synoptic observations. These are taken over an area, ideally at a number of altitudes over the entire globe, and at a number of times during the day. The resulting data are the basis of weather forecasting, whether from maps or from high-speed computers. *See* WEATHER (FORECASTING AND PREDICTION).

Operational observations. Such special observations are used to control some activity; for example, measurement of wind and temperature, visibility and cloud ceiling over an airport are necessary for efficient and safe operation of aircraft.

Research observations. Observations used in research may be any of the above or special ones taken with newly developed instruments.

Problems of instrument exposure. In addition to those generally desirable characteristics of instruments such as accuracy, reliability, and ease of maintenance, another requirement common to all meteorological instruments is care in exposure. A rain gage whose catchment area is 10 in. in diameter often represents the single rainfall measurement for an area more than 10 miles in diameter. Atmospheric properties are greatly affected by such factors as surface cover, proximity to trees or buildings, and elevation, to mention only a few. Because errors due to exposure cannot be completely eliminated, uniformity of exposure from station to station is essential. Often it is more difficult to obtain a representative sample than to secure high accuracy in the instrument itself.

The requirements for proper exposure are particularly stringent for climatological stations. The

records of a climate station might make it appear that the climate has slowly warmed when actually the environment of the station has changed because of the growth of a city around it. Likewise, it might appear that wind velocities are gradually becoming less—all caused by growth of trees in the vicinity. So important is the question of proper instrument exposure that the U.S. Weather Bureau has established a number of "bench-mark" stations carefully selected for likely constant conditions. Changes in a station record can then confidently be taken to be changes in the general climate and not changes in the environment.

Accessibility and conversion of data. Still another feature necessary in meteorological instruments is rapid access to the pertinent data, particularly for synoptic observations If many hours are required to convert a particular set of measurements into meaningful meteorological data, much of its value is lost. It arrives too late to be of much use in the next weather forecast. (For a case where this is a major problem, *see* METEOROLOGICAL USF OF SATELLITES.) The sheer volume of worldwide detailed observations is formidable.

Surface observational practices. Considerable information on the common use of instruments for surface observations of various elements of weather can be found in separate articles. For examples on temperature, *see* AIR TEMPERATURE; THERMOMETER; on atmospheric pressure, *see* AIR PRESSURE; BAROMETER; on humidity, *see* HUMIDITY; HYGROMETER; PSYCHROMETER; on wind, *see* WIND; WIND MEASUREMENT; on rain and other forms of precipitation, *see* PRECIPITATION (METEOROLOGY); PRECIPITATION GAGES.

Upper-air instruments. Modern meteorology began in the late 1930s with the advent of regular soundings of the atmosphere from many stations on a broad geographical scale. Isolated explorations of the atmosphere using kites were made as early as 1749 and with manned and free balloons during the ni[illegible]
ascents with [illegible]
graphs, whi[illegible]
and humidit[illegible]
three-dimensional picture of the weather
Air [illegible]
velopn[illegible]
the su[illegible]
radiowind system was much progress possible in gaining a three-dimensional understanding of the atmosphere. The radiosonde is a light-weight miniature weather station carried aloft by balloons with a radio link to the ground. A radiosonde-radiowind system consists of a balloon, the radiosonde, ground equipment to receive and record the signals from the radiosonde, and a radio theodolite which measures the elevation and azimuth angles of the balloon, used to obtain *measurements* of the upper winds.

Meteorological balloons. Used for lifting radiosondes to high altitudes, these balloons are made of neoprene rubber and must be of high quality. In order to reach altitudes of 100,000 ft, the rubber must stretch enough to enclose a volume 100 times greater than it had at the earth's surface. In order to reach even higher altitudes, the volume must double for roughly every 15,000 ft of additional altitude. This can be done by using large 10-kg balloons only partially inflated at the ground. They become fully extended at 30,000 ft and then stretch to their maximum volume as they ascend to burst ing altitudes. These larger balloons weigh much more than the payload they carry. One reaches a point of very rapidly diminishing returns at altitudes of 140,000–150,000 ft. Indeed, no further increase in altitude is possible even with zero payload [illegible] Routine ra[illegible]

[illegible] constructed from thin [illegible]

higher. These balloons have a fixed volume, in contrast to the variable volume of the rubber balloons To prevent bursting, they are either vented or are built strong enough to resist the slight super pressure inside. They can float at a fixed altitude for days. The vented balloon must be automatically ballasted to compensate for heating and cooling during the day and night. Balloons made of Mylar are strong enough to withstand the increased internal pressure caused by solar heating and therefore require no complex ballast system. These are ideal platforms for fixed-altitude radiosondes called transosondes because they make it possible to obtain upper-air observations over vast ocean areas. The maximum altitude of plastic balloons is about the same as that for large rubber balloons. In order to carry instruments to higher altitudes, rockets must be used. For additional information on this technique, *see* METEOROLOGICAL ROCKETS.

Data conversion and transmission. Radiosonde flight equipment, in general use by meteorological services of the different countries, can be divided into two basic systems: (1) those in which temperature and humidity sensors produce dimensional changes which are mechanically coded for transmission to the ground, and (2) those in which sensors produce electrical changes directly. *See* TELEMETERING.

1. Mechanical-change type. The thermometer in the first group is usually a bimetal element or a stretched wire. The humidity element is a hair or goldbeater's-skin hygrometer. The pressure element is an aneroid capsule. Several schemes for coding these mechanical motions into meaningful radio signals are used. In the Väisälä radiosonde, the sensors are coupled to variable capacitors which shift the carrier frequency of the radio transmitter. A wind-driven switch connects to the transmitter each sensor variable capacitor and two fixed capacitors in sequence. The two fixed capacitors, one large and the other small, are needed for in-flight calibrations. Another system in common use converts the mechanical motions into variable

time intervals. A clockwork or constant-speed electric motor turns a helix or two-turn flat spiral which forms one contact. Contact levers, mechanically coupled to the meteorological sensors, make contact with the helix or spiral each revolution. The time a contact is made during the revolution is a function of the position of the sensor contact. The beginning of each revolution is given by still another fixed reference contact. The relative position in time of each meteorological contact compared with the reference contact is a measure of the position of each meteorological element.

Photoengraving and photoetching techniques make possible substitution of a disk or drum containing many groups of Morse code signals for the spiral or helix. The radio signals from the radiosonde are then readable as regular Morse code signals. In still another type of radiosonde, the mechanical motion is used to move μ-metal slugs to vary the inductance of small coils. These are made part of an audio-frequency oscillator. Changes in the sensors cause changes in the audio tone transmitted to the ground. A precision oscillator or audio-frequency meter at the ground station is necessary to interpret the signals.

2. Direct electrical-change type. The radiosonde system in use by the U.S. Weather Bureau and military service employs a mechanical bellows-operated commutating switch and variable resistors for the temperature and humidity elements The thermistor temperature element is a rod of ceramic material which possesses a large negative temperature coefficient of resistance, roughly 4%/°C The resistance of the element varies from 20,000 ohms at +60°C to 2,000,000 ohms at −90°C. The diameter of the rods is made as small as practicable so that the elements will respond rapidly to changes in temperature. They are coated with a reflective paint or are shaded in a duct to reduce heating from solar radiation.

The humidity element consists of a polystyrene strip provided with two electrodes on the edges. The space between the electrodes is coated with a weak solution of lithium chloride held in a binder of polyvinyl acetate and alcohol. The resistance of the strip varies with changes in vapor pressure and temperature. During flight, the pressure-actuated commutating switch operates a relay. In the normally closed position, the relay connects the temperature element into an electronic blocking-oscillator measuring circuit. In the alternate position, the relay connects the humidity element into the circuit. The value of pressure is determined by the number of times the humidity, temperature, and a fixed-value resistor have been connected into the circuit, whereas humidity and temperature are determined from the pulse repetition frequency of the blocking oscillator.

A notable feature of the radiosonde in use in the United States is the simple linear frequency vs. electrical conductance characteristic of the blocking-oscillator circuit. The frequencies of the meteorological sensor signals can be compared with the frequency generated by a known calibration resistance included in the sensor switching sequence. Because of this feature, the manufacture and calibration is greatly simplified. Careful matching of temperature thermistor elements and humidity elements to a specific blocking oscillator is unnecessary. Within reasonable limits any temperature element will work with any blocking oscillator, and a simple preflight calibration establishes the ratio for a particular combination. The pulses from the blocking oscillator modulate a 1680-megacycle/sec cavity oscillator transmitter to provide the radio link to the ground station.

The ground equipment not only includes a radio receiver and recording audio-frequency meter for reception of temperature, pressure, and humidity signals, but also an automatic tracking antenna. The latter, a parabolic dish about 10 ft in diameter, is capable of measuring the elevation and azimuth angle to the air-borne radiosonde to a few hundredths of 1° This radiotheodolite angle data and the height of the balloon available from the radiosonde data allow the computation of upper winds to maximum altitude, even though the balloon is hidden by cloud In order to improve the accuracy, particularly at high altitude and long range, a ranging unit is sometimes used.

For the ranging unit a tiny radio receiver is added to the balloon-borne equipment to make a transponder This permits measurement of the round-trip time for a radio signal, from the ground station to the balloon and back again, to give the slant range The slant range multiplied by the cosine of the elevation angle gives the horizontal distance to the point under the balloon, whereas when the height of the balloon is used to obtain the distance, the product with the cotangent of the elevation angle is required. Because at low angles of elevation for a given error in angle measurement, the cosine varies much less than the cotangent of the angle, the accuracy of the wind measurement at high altitude and long range is considerably enhanced by use of the ranging device.

Ground evaluation and dissemination. The data from a typical radiosonde receiver flight is evaluated in about 1 hour, coded for rapid transmission by teletype, and transmitted to weather stations throughout the world. *See* WEATHER (FORECASTING AND PREDICTION).

Radar wind observation. Radar is also used to obtain data on upper winds. The radar beam is reflected by a passive target which is part of, or attached to, the balloon. The balloon itself may be coated with a thin metallic film to make it a conductor. In order to obtain a strong return signal, a corner reflector constructed of aluminum foil with a light framework is used. A corner reflector tends to reflect the incident electromagnetic energy back toward the radar antenna from whence it came, much as a billiard ball moving from one corner of a billiard table toward the opposite corner tends to return to the starting corner. Because the radar has excellent angular and range capability, the wind

determination by radar is usually quite good. When an active transponder is used instead of a passive reflector, wind measurements at great range and altitude are possible.

Radar storm tracking. Radar is of even greater utility for storm detection. Strictly speaking, radar detects the precipitation associated with the storms. It presents a graphic picture of the precipitation area in the three dimensions together with all the changes that occur. The radar echo information is usually presented on an oscilloscope as a horizontal map with the location of the radar set in the center. This data appears on the plan position indicator (PPI scope, or PPI). A vertical picture of the precipitation area at a fixed azimuth angle is presented on the range height indicator (RHI scope). The antenna on a tracking radar is automatically locked on to the balloon target, whereas on a search radar used for storm detection the antenna rotates continuously to illuminate a horizontal sheet for the PPI presentation and moves up and down in angle of elevation for the RHI picture.

In addition to providing the location of a precipitation area, the strength of the echo signal is related to intensity of the precipitation. In thunderstorms, the turbulent cells are usually associated with the heaviest precipitation. Radar-equipped aircraft are able to detect and avoid these dangerous parts of a storm, hence the name weather-avoidance radar. There is no simple relationship between the strength of the radar echo and the amount of precipitation in the cloud (rainfall if it all falls to the ground). The strength of the echo signal is related, among other parameters, to the number of drops in a given volume and the sixth power of the drop diameters. The amount of liquid water is, however, proportional to the number of drops and the third power of the drop diameters. Hence, to obtain the amount of rainfall, the drop-size distribution must be known independently. Attenuation of the radar signal by other rain between the target in question and the radar set complicates the determination further. The power scattered by a given drop size is greater for short wavelengths, typically 3-cm wavelength, than for larger wavelengths, typically 10 cm. Thus it is possible to detect light rainfall with short-wavelength radar but further penetration into a storm requires 10-cm radar. The optimum wavelength for weather radar is near 5–7 cm. Despite these seeming shortcomings, weather radar has been of enormous benefit, especially through forecasting. Violent thunderstorms, tornadoes, and hurricanes have been located and tracked on a routine basis by the weather services of many countries. [V.E.S.]

Meteorological optics

A branch of atmospheric physics or physical meteorology in which optical phenomena occurring in the atmosphere are described and explained. Meteorological optics can be divided into two main parts. In the first part, the atmosphere is considered as a continuous medium in which the speed of propagation of electromagnetic waves (light) varies with the density. In the second part, attention is given to the presence of particulate matter in the atmosphere, such as air molecules, dust and haze particles, cloud and raindrops, and ice crys- [illegible] from the inter- [illegible] part [illegible] t of [illegible] atmospheric properties.

Variations of refractive index. The first part of meteorological optics discusses optical phenomena in the atmosphere caused by the variation of the speed of light propagation, or of the refractive index, as a result of changes in the air density, such as (1) astronomical and terrestrial refraction, for instance the bending of light rays as they penetrate the denser layers of the lower atmosphere, causing the apparent increase of the angular elevation of celestial or terrestrial objects above the horizon; (2) mirages, in which the bending of the light rays is due to abnormal vertical distribution of air density (*see* MIRAGE); (3) scintillation, a condition of rapid changes of images of stars or other distant bright objects caused by rapid variations in the air density; (4) green flash, or the green color of the last visible segment of the setting sun, owing to the dispersion and attenuation of the horizontal rays.

Interaction with particulate matter. The interaction of light with particulate matter can be twofold. The radiant energy, received by a molecule of air or by a particle floating in the air, can be changed totally or in part into another type of energy such as heat or photochemical energy. This process of energy change is called absorption, and the amount of energy transformed represents a loss of radiant energy. In the visible range of the spectrum of electromagnetic waves, however, the atmospheric gases, except ozone, do not absorb light. Much more important is the second type of interaction of light with the particulate matter. In this process, called scattering, the energy received is completely reradiated in all directions around the scattering center, so that no part of the radiant energy is transformed into another type of energy and thus lost. In the second part of meteorological optics, the different types of scattering as well as the important consequences of light scattering are discussed.

Light scattering by air molecules. This type of scattering is governed by a law, first formulated by J. W. Strutt, Lord Rayleigh (1871). It applies to any shape of particle, the dimensions of which are negligible with respect to the wavelength of the incident light. The intensity of light scattered by such a particle is inversely proportional to the fourth power of the wavelength, being more intense in the violet than in the red part of the spectrum. The scattered light is polarized; that is, when observed through an analyzer such as a Nicol prism or Polaroid plate, it shows variations in the intensity while the analyzer is rotated around its axis

(two maxima and two minima when rotated by 360°). The blue color of the sky and the polarization of skylight in a pure atmosphere containing only air molecules can be explained as consequences of light scattering according to the Rayleigh law. Another important consequence of light scattering is the attenuation or extinction of light due to scattering. The radiant energy which is scattered in all directions by the scattering center can be noticed as a loss of the radiation in the direction of the incident light. The amount of attenuated radiant energy is also inversely proportional to the fourth power of the wavelength. Therefore the light rays of a long path length will contain less violet light than red. The loss of the violet and blue light is particularly noticeable in the change of color of the setting sun, from white to red.

Mie scattering. The scattering of light by particles of size comparable to or larger than the wavelength of the incident light is characterized by greater intensity of the scattered light in the forward direction than in the backward direction. For a dielectric sphere the law of scattering was developed by G. Mie and is sometimes called Mie scattering. The light scattered by dust and haze particles in the air varies with the wavelength much less than in the case of molecular scattering, and may even be independent of the wavelength. Thus the intensity of the scattered light is much larger in the long wavelength compared with that for molecular scattering. The presence of dust and haze particles in the atmosphere can be noted by the change of the deep blue color of the sky in a pure atmosphere into a whitish blue color of the sky in a turbid atmosphere, with a bright white region around the sun, called the aureole. The attenuation of light due to the scattering by dust and haze particles is very nearly independent of the wavelength, so that the setting sun, seen through a dense layer of haze or mist, does not change its white color. *See* SCATTERING (ELECTROMAGNETIC RADIATION).

Large-particle optics. The interaction of light with large particles, such as water droplets, in fog, cloud or rain, is governed by the laws of diffraction and geometrical optics. Interference of rays passing very close to the edge of the drops gives rise to a series of rings of the maximum and minimum intensity called corona and visible as a system of rings of different colors around the illuminating bodies, sun or moon. A similar interference effect occurs in the backward direction, around the antisolar point, and becomes noticeable as a system of colored rings around the shadow cast by the sun on a layer of cloud or fog, the anticorona or glory. The rays which are reflected once or twice inside the sphere cause the colored phenomena of the primary and secondary rainbow (*see* RAINBOW).

Ice crystals appear in the atmosphere in the form of hexagonal prisms. Part of the light incident on the ice crystal is refracted as by a prism with the sides forming an angle of 60 or 90°. As a consequence of such refraction, two luminous rings appear around the sun or moon, as part of a complex phenomenon called the halo (*see* HALO). On both sides of the sun two bright spots, called parhelia, occasionally appear provided that the hexagonal prisms have a prevailing vertical orientation of their axes (*see* SUNDOG).

Visibility and related features. Closely connected with the scattering of light in the atmosphere is the problem of visibility. The light scattered into the observer's eye by particles situated along the line of sight to a distant object diminishes the contrast of the object and its background, thus decreasing its visibility.

By a combination of the effects of terrestrial refraction and of light scattering the twilight phenomena can be explained, particularly the different colors of the sky, the twilight arcs, seen above the sun horizon as well as in the opposite direction, above the rising earth shadow.

Applications to related problems. Several phenomena, some of which are enumerated above, can be used for indirect measurement of certain physical quantities. For example, air density may be judged by measurement of the amount of lifting of distant objects due to the anomalous refraction of horizontal rays. Similarly, the measurements of the skylight intensity and its polarization, as well as of the attenuation of the sun radiation, can be used for the determination of atmospheric turbidity. The measurement of the angular distances from the sun of different rings in the corona can be used for estimating drop sizes in a cloud. For the same purpose similar measurements can be applied to the anticorona or to the rainbow. The measurements of the attenuation of sun radiation, combined with the measurements of the intensity of skylight in the ultraviolet, was successfully used for the determination of the total amount and vertical distribution of ozone in the upper atmosphere.

Several technical problems can be solved by direct application of the results of meteorological optics, such as the amount of illumination of objects on the ground or high in the atmosphere and the conditions of their visibility under different atmospheric conditions. [Z.S.]

Bibliography: S. Fluegge (ed.), *Handbuch der Physik*, vol. 48, 1957; W. J. Humphreys, *Physics of the Air*, 3d ed., 1940; T. F. Malone (ed.), *Compendium of Meteorology*, 1951; H. C. van de Hulst, *Light Scattering by Small Particles*, 1957.

Meteorological rockets

Small rockets used to extend observations of atmospheric character above currently feasible limits for balloon-borne observing and telemetering instruments. Present balloon-borne rawinsondes (radiosondes with wind-observing apparatus) permit upper air observations to about 100,000 ft. It is doubtful that they can be extended appreciably above this altitude on a routine basis. Recent research indicates that the atmosphere in the zone



The Arcas rocket. This small (77 lb) rocket is being developed to carry 6.5 lb of instrumentation and a 3-lb parachute to an altitude of 200,000 ft at which payload is separated from the rocket motor. Measurements are obtained during descent of instrumentation on the parachute. (*Atlantic Research Corporation*)

from 70,000 to 300,000 ft may have an important interaction with the atmosphere below 70,000 ft. The meager information presently available on winds at altitudes above 100,000 ft indicates that they are often very strong and undergo a systematic reversal in the general directional flow from winter to summer. Some recent rocket observations also show large diurnal variations in the wind at about 200,000 ft. Many more rocket observations are needed for full description and understanding of these and other significant features in the atmosphere, and to provide needed information on the environment encountered by high-altitude vehicles.

Because of the continual changes in the atmosphere, meteorologists require frequent soundings simultaneously at several locations. Therefore, any successful rocket program for research must use low-cost components. The first low-cost rocket to be used for meteorological observations above 100,000 ft is the Loki. This small, solid-propellant rocket motor propels a projectile to maximum altitude at which radar chaff or a metallized parachute is ejected. The atmospheric winds are observed and computed by using a radar to track the target during its descent. Observations have been obtained to 280,000 ft. Although only a limited number of observations have been made with the Loki system, plans are being formulated to initiate a regular observing program at several points in North America.

Although some development is under way, at present there is no low-cost rocketsonde system which will permit the measurement of temperature and pressure (or density) as well as winds. The most probable configuration of the first such system is a single-stage, solid-propellant rocket which will carry the instrument package to about 200,000 ft. Measurements will be obtained as the instrument is lowered by parachute. A transmitter in the instrument package will be tracked to obtain the winds. Temperature, measured by a sensor such as a very small thermistor or fine resistance wire, will be telemetered by the transmitter.

Additional instrumentation and new techniques must still be developed to measure inexpensively pressure (or density); atmospheric constituents, particularly water vapor, ozone, and carbon dioxide; and to extend observations to about 300,000 ft. *See* METEOROLOGICAL INSTRUMENTATION. [D S.J]

Meteorological use of satellites

Artificial satellites have two main attributes of value for meteorology: (1) they can be equipped to provide world-wide meteorological observations as they scan the world in one day or less; and (2) they can be used to observe the earth from outside and provide data which have rarely been obtained. For example, a satellite television camera can transmit pictures showing cloud amount and cloud type from all over the sunlit world. In the Pacific Ocean area, observations are sparse now; there the televised cloud data of storms and their development may assist in forecasting for the United States. In the tropical Atlantic, hurricanes can be detected in their initial breeding grounds and tracked, often far from any earthbound observer. In the Southern Hemisphere, in polar regions, and in other inaccessible areas cloud measurements could ultimately provide important weather information including, at times, wind data. Even in areas where there is a dense network of ground observation stations, a simultaneous televised picture of all clouds would be a valuable supplement.

The satellite can be used to provide measurements of the heat budget of the atmosphere. The sun irradiates the earth; part of this solar energy is reflected, unused, to space and part is absorbed by the earth's atmosphere and surface. Moreover, the earth radiates energy to space day and night. The satellite may be equipped to measure these solar and terrestrial energy streams, which vary geographically and in time with the cloudiness and composition of the atmosphere and with the changing character of the earth's surface. This important information could be related to variations in the general circulation of the atmosphere, which is basically driven by those energy streams.

Many other meteorological measurements are being studied for adaptation to satellites. One of these, cloud-cover measurements at night, could ultimately supplement daytime observations by television. The water vapor of the atmosphere is transparent to infrared radiation from the earth in the spectral region of about 10 μ wavelength, called the atmospheric water vapor window. Because moderately thick clouds of liquid or ice droplets are not transparent at this wavelength, an infrared sensing device in a satellite could distinguish

between cloud-covered and clear areas. In the tropics the infrared sensing device could aid meteorologists to measure the height of cloud tops at night as well as day. This is so because emission of infrared radiation by a thick cloud depends only on its temperature, and in the tropics temperature often varies in a known relation to height above the earth.

Satellite radars measuring precipitation all over the world would aid studies of the energy flux in the atmosphere, as well as synoptic research. Some radars may also contribute to cloud data.

Measurements in the water-vapor absorption bands, such as the 6-μ band, may be used to learn about the world-wide water vapor distribution, the temperature near the tropopause, or both. Measurements in the 15-μ CO_2 band will probably indicate the temperature distribution in the stratosphere. The emission in the 9.6-μ ozone band may reveal the distribution of atmospheric ozone in the stratosphere; measurements of *reflected* solar ultraviolet light, partially absorbed by ozone, may supplement the infrared ozone measurements. The optical properties of the earth and its atmosphere, such as its color and dust content, should be deducible from the spectrum of reflected solar radiation.

In summary, the satellite can provide world-wide observations of types of data now only sparsely observed, and also of types of data not now observed at all. [S.FR.]

Meteorology

The science concerned with the atmosphere and its phenomena.

Meteorological science is primarily observational; its data are generally "given." The meteorologist observes the atmosphere—its temperature, density, winds, clouds, precipitation, and other characteristics—and aims to account for its observed structure and evolution (weather, in part) in terms of external influence and the basic laws of physics. Empirical relations between observed variables, as those between the patterns of wind and weather, are developed to pose more effectively the problems to be investigated and explained and to provide essential material for the application of the science. Weather forecasting serves as an example of such application because theory still remains insufficiently developed to provide more certain applications. *See* WEATHER (FORECASTING AND PREDICTION). Very little controlled experiment has been made on the atmosphere but more is probable.

This background article has a threefold organization. The first portion presents a summary of the general physics of the air; this has been the principal approach and basis for meteorological science and its applications, such as to weather phenomena (the condition of the atmosphere at any time and place) and climate (a composite generalization of weather conditions throughout the year, *see* CLIMATOLOGY). A second portion, synoptic meteorology, discusses the character of the atmosphere on the basis of simultaneous observations over large areas. Concurrently, and at an accelerating pace, dynamic principles (thermodynamic and hydrodynamic) are being applied to meteorological investigations. This study of naturally produced motions in the atmosphere is forming much of the scientific basis for modern weather forecasting and physical climatology. Hence, the third portion of this article deals with dynamical meteorology.

GENERAL PHYSICS OF THE AIR

The components of dry air, excluding ozone, and their relative volumes (mol fractions of gases) up to a height of at least 50 km are given in Table 1. Some of the rarer gases, such as CO_2, are continually entering and leaving the atmosphere through the earth's surface, and the fractions quoted are thus mean values. The effective molecular weight of the mixture is 28.966 and its equation of state, to 1 in 10^4, is $p = R\rho T$ where p is pressure, ρ density, T absolute temperature and R, the specific gas constant, is 2.8704×10^6 erg g^{-1} °K^{-1}. *See* AIR; ATMOSPHERE.

Above 50 km, O_2 becomes progressively dissociated to O, which is probably dominant above about 120 km. N_2 probably dissociates at appreciably higher levels. There is no good evidence for diffusive gravitational separation of the lighter from the heavier gases below 100 km—nor indeed is this likely.

Water vapor and ozone are highly variable additional components of air. The fraction of the former commonly decreases rapidly with height, from about 10^{-2} at sea level to 10^{-6} or 10^{-7} at about 16 km, with dissociation at much higher levels. Ozone, the product of photochemical action in sunlight at high levels, has a maximum fraction ($\sim 10^{-5}$) at about 25 km. The importance of these two constituents is far greater than their fractions might suggest because (1) both are radiatively active, water vapor and ozone in the infrared of terrestrial emission, ozone in the near ultraviolet of solar emission; (2) water vapor condenses to the liquid or solid, giving cloud, which reflects upward a major part of solar radiation incident upon it, and in condensing releases a large latent heat to the air. Carbon

Table 1. Components of dry air

Gas	Symbol	% vol
Nitrogen	N_2	78.09
Oxygen	O_2	20.95
Argon	Ar	0.93
Carbon dioxide	CO_2	0.03
Neon	Ne	1.8×10^{-3}
Helium	He	5.2×10^{-4}
Krypton	Kr	1×10^{-4}
Hydrogen	H_2	5×10^{-5}
Xenon	Xe	8×10^{-6}
Nitrous oxide	N_2O	3.5×10^{-5}
Radon	Rn	6×10^{-18}
Methane	CH_4	1.5×10^{-4}

dioxide, CO_2, is the only other constituent with strong infrared activity.

Thermal structure. It is convenient to divide the atmosphere into a number of layers on the basis of its thermal structure. The first layer (*Fig. 1*) is the troposphere, in which temperature on average decreases with height 6C°/km (the "lapse rate") everywhere except near the winter pole. The troposphere is about 16 km deep in the tropics and about 10 km deep, with substantial variations, in higher latitudes. At any one level in the layer, the mean temperature decreases from equator to pole, by about 40°C in winter and 25°C in summer.

The second layer is the stratosphere in which the temperature varies at first very little with height and then increases to near the surface temperature at about 50 km. In this layer temperature increases polewards, except quite near the winter pole. The boundary between troposphere and stratosphere is termed the tropopause, which may drop abruptly in altitude at about 30° latitude and again in higher latitudes.

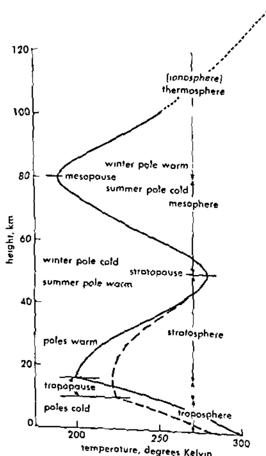


Fig. 1. The vertical structure and nomenclature of atmospheric layers in relation to temperature. The values of temperature shown are approximate; it is the form of the temperature variation with height which determines the nomenclature. The two curves shown for the troposphere and stratosphere refer to lower (full line) and higher (dashed line) latitude conditions. The predominance of the meridional temperature gradient at various levels is shown by entries "poles cold," "poles warm." A change of sense of meridional gradient appears to coincide with a change of vertical gradient (pause) only at the tropopause.

From 50 km to 80 km the temperature again decreases with height to an absolute minimum of about −80°C. This layer has been called the mesosphere and its lower and upper boundary the stratopause and mesopause respectively. See MESOSPHERE; STRATOSPHERE; TROPOSPHERE.

Above 80 km the temperature again increases with height, at a rather uncertain rate. The layer is strongly ionized by solar ultraviolet and other radiations and is variously termed the ionosphere or thermosphere. See IONOSPHERE.

Radiation and thermal structure. During the passage of solar radiation downward through the atmosphere, the following, in broad outline, takes place. The ultraviolet and shorter waves are absorbed by the thermosphere, more above than below, so that temperature may be expected to fall along the path of the beam. Entering the mesosphere, O_3 (ozone) is encountered, and new absorption of energy takes place in the near ultraviolet. The increase of ozone concentration along the path outweighs the depletion of the beam energy by O_3 at upper levels so that the temperature increases along this portion of the path. Beneath the stratopause, however, depletion in the wavelengths concerned has become large and the O_3 concentration itself ultimately falls so that temperature decreases along the path. There is little absorption in the troposphere—the slight near-infrared absorption by water vapor is practically uniform along the path because of the increasing concentration of vapor—but there is substantial back-scatter by air molecules and by clouds. The earth's surface, except snow, absorbs the incident solar radiation strongly and this leads to the heating of the atmosphere from below by various kinds of convection and to a lapse of temperature in the troposphere. See INSOLATION; RADIATION, TERRESTRIAL.

Because the atmosphere and underlying surface maintain their temperature over the years, the terrestrial emission of radiation to space must nearly balance the absorption of solar energy. The balance is achieved globally but not locally, the winds providing the necessary transport of heat from the regions of net absorption to the regions of net emission. While the broad form of the temperature profile of Fig. 1 is determined by the pattern of solar absorption, the actual temperatures are also due to the terrestrial emission and wind transport. See HEAT BALANCE, TERRESTRIAL ATMOSPHERIC.

Pressure and wind. The pressure at any point in the atmosphere is the weight of a column of air of unit cross section above that point. Therefore pressure decreases with height. At any level it decreases more rapidly where the air is cold and dense than where it is warm and light. The pressure at sea level is generally a little over 1 kg/cm²,

varying by a few per cent in space and time because of the inflow of air over some regions and its outflow over others. The horizontal pressure pattern necessarily changes with height when, characteristically, the temperature varies horizontally.

In an unheated, nonrotating atmosphere, air would flow horizontally from high pressure to low to remove the pressure difference. Other forces arise in the actual atmosphere to provide a flow which is more nearly along, than perpendicular to, the isobars (lines on a map connecting areas of equal barometric pressure). Therefore pressure patterns persist for days, with gradual modification, and are translated with speeds often comparable with the winds blowing in them. The effect of the earth's rotation on the wind is nearly always dominant, except very near the equator. If it precisely balances the pressure force, the wind blows with speed V along the isobars (for an observer with his back to the wind, low pressure is to the left in the northern hemisphere)

where $$V = \frac{(\partial p/\partial n)}{2\omega\rho \sin\phi}$$

($\partial p/\partial n$ is the horizontal pressure gradient, ω the angular velocity of the earth, ρ the air density and

Fig 2 *The pattern of mean zonal (east-west) wind speed averaged over all longitudes as a function of latitude, height, and season (after Y. Mintz). Height, greatly enlarged relative to earth radius, is shown on a linear pressure (α mass) scale with geometrical equivalent given at upper left. The mean zonal wind (westerly positive, easterly negative—stippled parts) has the same value along any one line and is shown in meters per second (1 m/sec ≅ 2 knots) on the line. Regions of easterlies are shown shaded. Note the subtropical westerly jet in the high troposphere at about 30° latitude (or more in northern summer). This jet is not to be confused with the polar front jet of higher latitudes; the latter is migratory and so does not appear in the mean.*

ϕ the latitude). This wind V is called the geostrophic wind and is a good approximation to the actual except near the surface, where friction always intervenes. Since, as seen above, the pressure gradient generally varies with height, so do both V and the actual wind. The change of V with height is proportional to the horizontal temperature gradient, low temperature being to the left of its direction of movement in the northern hemisphere. *See* AIR PRESSURE, WIND.

General atmospheric circulation. The mean wind in nearly all parts of the atmosphere is predominantly along the latitude circles. The pattern of this zonal motion in latitude and height is shown in Fig. 2 and is the basis of the general circulation of the atmosphere. Surface easterlies (trade winds) are found in the tropics, surface westerlies in middle latitudes, and surface easterlies again near the poles. But everywhere in the troposphere except near the equator, winds from the west increase in strength with height; in the tropics and polar regions [illegible]

winds is consistent with the meridional gradient of temperature in the troposphere and its reversal in the stratosphere, if the winds are quasi-geostrophic. *See* GEOSTROPHIC WIND.

The absolute maximum of zonal wind in Fig. 2 is found in the upper troposphere at about latitude 30° and is the core of the subtropical jet stream. There are other substantial mean zonal motions at higher levels than can be shown in Fig. 2 but they are not well explored. *See* JET STREAM.

The general circulation is maintained against frictional dissipation by a "boiler-condenser" arrangement provided by the heat absorbed, directly or indirectly, from the sun and that emitted to space by the earth. The wind thereby generated transports heat from source to sink to maintain thermal balance, disturbances on the mean motion being important in the process.

Water vapor, cloud, and precipitation. The motion of the air is rarely quite horizontal; it is commonly moving upward or downward at a few centimeters per second over large areas, while locally the vertical velocity may be several meters per second. Air which rises is cooled by expansion into lower pressure and that which descends is correspondingly compressed and warmed. Since air is always more or less moist a sufficient rise will cause the temperature to drop to the saturation point and cloud will then form on nuclei present in the air. Large-scale ascent leads to extensive sheets of cloud called stratus; cirrostratus at high, ice-forming levels, altostratus at medium levels, nimbostratus at lower levels. Local, strong ascent produces heap clouds (cumulus or cumulonimbus). If the cloud base is sufficiently warm or the cloud depth sufficiently great, or both, the condensed water or ice in the cloud falls out as rain, snow, or other precipitation. Certain lenticular clouds are due to local ascent and descent brought about by hills, and these clouds do not move with the wind. Cloud formation is practically confined to the troposphere because this alone contains sufficient water vapor and sufficiently sustained vertical motion. *See* CLOUD; CLOUD PHYSICS; PRECIPITATION (METEOROLOGY).

Atmospheric disturbances. The mean motion, described under the general circulation of the atmosphere, is disturbed by many patterns of motion of varying scale. A chart of the flow well above the surface over the northern hemisphere will generally show, on any day, large meanderings of the air poleward and equatorward superposed on a general westerly flow. These are the long waves of the westerlies, with a few thousand kilometers between the turning points. They are generally associated with a jet stream in the upper troposphere. Another class of large-scale disturbances, seasonal in nature and most apparent in the lower troposphere, is designated the monsoons; they are particularly developed in relation to the south Asian cyclone of summer and the Siberian anticyclone of winter. *See* AIR WAVES, UPPER SYNOPTIC; MONSOON.

Proceeding downward in scale are the traveling depressions (cyclones) and anticyclones, ridges and troughs of extratropical latitudes, 1000–2000 km in horizontal extent, and to the somewhat shorter waves in the easterlies (trades). Next in size is the tropical cyclone (~100 km)—the hurricane or typhoon, according to location—which is mainly confined to the oceans and eastern seaboards of continents. Still smaller (~1 km) is the tornado, mainly confined to land and associated with cumulonimbus cloud, and the waterspout associated with similar cloud over the sea. The smallest "revolving storm" is the dust devil (~10 m), which occurs immediately above a hot, dry surface in a very light general wind. *See* ATMOSPHERIC HIGH; ATMOSPHERIC LOW; CYCLONE; FRONT; STORM; TORNADO.

In addition to the above well-defined patterns of flow there are randomly distributed fluctuations of flow over several orders of magnitude of scale from about 1 cm upward, particularly evident in the bottom kilometer of the atmosphere and in and around cumulus. They are related to friction, convection, and wind shear and are important agents of transfer of heat, matter (water vapors, dust, ozone), and momentum.

Air masses and fronts. The horizontal variations of air temperature referred to in previous paragraphs are commonly concentrated into narrow zones, or even discontinuities, with low gradients in intervening areas. These narrow zones or discontinuities are called fronts, and the air in a region of small temperature gradient is called an air mass. If a front moves so that a warm air mass replaces a cold air mass we speak of a warm front, and conversely of a cold front. Fronts may also be stationary. They slope upward at a slope ratio of

about 1 in 100, with the cold air as a wedge beneath the warm air.

Most extratropical depressions first appear as dents or waves on a frontal surface, warm air rising slowly over an extended area in the forward part of the wave. This results in stratiform cloud, and a stronger updraft immediately ahead of the rear part of the wave yields cumulonimbus.

The tropopause is higher and its temperatures lower above a warm air mass than above a neighboring cold air mass, and a break, or offset, appears at the frontal boundary with a jet stream in the warm air mass near the break in the tropopause. *See* AIR MASS; TROPOPAUSE.

Optical phenomena. Scattering, refraction, and reflection of light by the air or by particulate matter (dust, cloud particles, or rain) in the air give rise to a variety of optical phenomena. *See* METEOROLOGICAL OPTICS.

Electrical properties. The atmosphere and the underlying earth are like a leaky electrical capacitor. Positive charge is separated vertically from negative charge in thunderclouds and some other areas of disturbed weather; the net result is that the earth's surface is left in fine-weather areas with an average negative charge σ of 2.7×10^{-4} esu/cm² to which corresponds a vertical field F_0 ($= 4\pi\sigma$) at the surface of 100 volts/m. The other "plate" of the capacitor is the highly conducting upper atmosphere, and the leak arises from the small conductivity of the air between the plates, produced by ionization by radioactive matter in the soil and air, and by cosmic radiation, the conductivity increases with height. The resistance R of a 1-cm² column of atmosphere is about 10^{21} ohms and an air-earth current $i = V_\infty/R$ (V_∞ being the potential of the upper conducting layer) of about 2×10^{-16} amp/cm² flows as a discharge current. V_∞ is thus about 2×10^5 volts above earth. The increase of conductivity with height implies a proportionate decrease of the field with height, and this in turn implies a small positive ionic space charge in the air.

The air-earth current would discharge the capacitor in about ½ hour if V_∞ (or σ) were not maintained by the charge separation in thunderclouds. This separation is more than adequate, the excess providing lightning flashes within the cloud—a shorting of the generator. *See* ATMOSPHERIC ELECTRICITY. [P.A.S.]

SYNOPTIC METEOROLOGY

This branch of meteorology comprises the knowledge of atmospheric phenomena connected with the weather, applied mainly in weather forecasting, and based on data acquired by the synoptic method. This method involves the study of weather processes through representations of atmospheric states determined by synchronous observations at a network of stations, most of which are at least 10 km apart. By international agreement, the data taken from the earth's surface and aloft at certain international hours of observation are inserted on weather maps, upper-air maps, vertical cross sections, time sections, and sounding diagrams with international symbols according to fixed rules. These crude representations are then analyzed and critically evaluated in accordance with the knowledge of existing structure models in the (lower) atmosphere, in order to ascertain the best approximation to a three-dimensional image of the true atmospheric state at the hour of observation. From one such representation, or a series of them, future atmospheric states are then derived with the aid of empirical knowledge of their behavior and by application of the theoretical results of dynamic meteorology.

General atmospheric circulation. Figure 3 shows some of the main constituents of the average state in the bottom layer of the atmosphere as to flow and pressure patterns, main air masses and fronts. The illustration also indicates how these large-scale mechanisms form part of a general atmospheric circulation with trade winds, monsoons, high-reaching middle-latitude westerlies, and shallow polar easterlies (see the vertical cross section of Fig. 2).

The planetary high-pressure belt at 30°N and S is split into subtropical high-pressure cells mainly as a result of the joint dynamic and thermodynamic effect of the great continents and mountain ridges of the earth, especially the Cordilleran highlands of South America. These cells again determine the formation and average position of the different polar fronts and air masses at the earth's surface, which are so manifest in the Southern Hemisphere.

Other general structures in the atmospheric circulation of importance to weather are the quasi-horizontal tropopause layers at different heights within different air masses and the tropical fronts, in the United States called Zones of Intertropical Convergence (ITC). The former generally mark the top of any considerable convection or upglide motion and of clouds in the atmosphere, but represent no store of potential energy. The latter form at the meeting of two opposite trade wind systems in the doldrums and have functions partly similar to those of the polar fronts, although they are much weaker and less distinct.

Air masses and front↔jet systems. Air masses may be classified in two distinctly different ways: thermodynamically and geographically.

Thermodynamic classification. The first classification is based on their recent path and life history, for example, the two opposite cases: the air being warmer or colder than the earth's surface, resulting in warm mass and cold mass. The warm mass, usually flowing poleward, is much warmer than the seasonal normal of the region, at least aloft. With the cooling from below, it gradually acquires a stable stratification, which damps turbulence and vertical mixing. Thus, the wind is relatively steady, the visibility low, and advection fog or stratus clouds often form within it, at least at sea, sometimes even yielding drizzle. This air

is most typically found within (upper) warm highs (see Fig 10), where the air is subsiding. These highs bring periods of steady and rather dry weather —very warm in summer on land—to middle and higher latitudes. *See* WIND.

The cold mass, mostly flowing equatorward, is colder than normal, at least aloft. When heated from below it rapidly acquires an unstable stratification which favors turbulence and vertical mixing or convection. Thus, the wind is gusty, visibility is good, and usually where the air is moist enough convective clouds form: cumulus → cumulonimbus with showers → thunderstorms and even hail. This air mass is most typically found within (upper) cold lows (Fig. 10), where if there is sufficient moisture the instability, general convergence, and

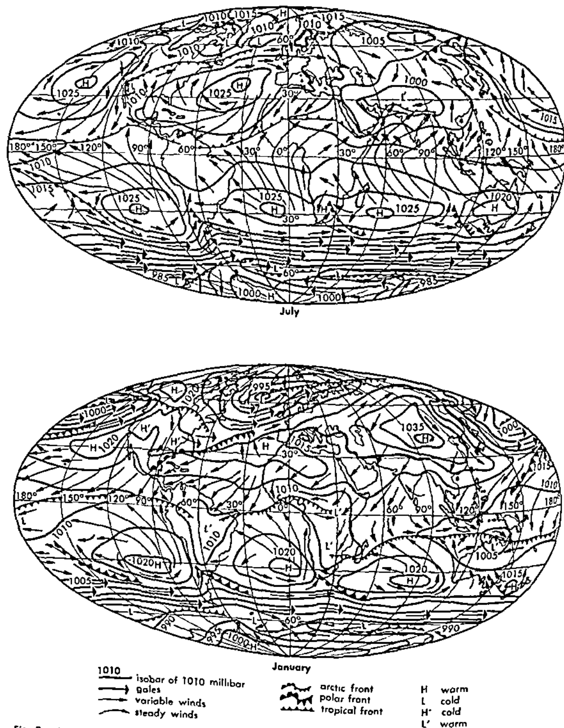


Fig. 3. Air masses and fronts as links in the general atmospheric circulation.

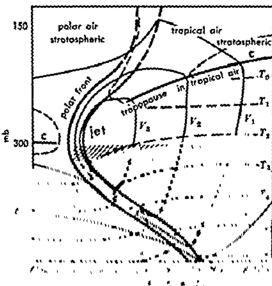


Fig 4 ... ross section of a polar-front ↔ jet system.

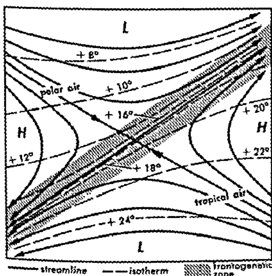


Fig. 5. Frontogenesis in a field of deformation.

lifting tendency of the air may favor the formation even of nonfrontal rain areas, which are discussed later. Therefore, such a low usually brings a period of wet and stormy weather to the southern part of middle latitudes, and in winter to the subtropics. However, at night over land, even this air may be stabilized so efficiently that radiation fog occurs within it.

Geographic classification. The second classification is based on the geographical origin or position of the air and the values of characteristic properties (such as temperature and specific humidity).

Tropical air occupies all the space between the polar-front ↔ jet systems of both hemispheres; aloft it may reach far into the polar region. At the midtroposphere this air is 10–20°C warmer and much more humid than the polar air at the same level and latitude. At low levels, the tropical air emerges from the quasi-stationary subtropical highs, reaching middle latitudes from the southwest, particularly within the warm sectors of migrating cyclones (Fig. 7), as a mild or warm, moist, and hazy air current (the stable warm mass). Aloft, it appears in high latitudes within the warm highs (Fig. 10). The tropical air also flows equatorward and westward within the northern and southern trade wind systems. When approaching the tropical fronts (Fig. 3) it appears, at sea, as an unstable cold mass with intense convection and shower activity on both sides of the front: the equatorial rains Over land, for example in North Africa, only the southwestern monsoon (that is the southeastern trade wind that has invaded the other hemisphere) is moist enough to produce much convective clouds and rain.

Polar air is found on the polar side of the polar fronts, as shown in Fig 3. Below, it emerges in winter from continental subpolar highs between 40° and 60°N, and in summer from the polar basin. The polar masses, and still more so the arctic and antarctic air masses, are mainly characterized by low temperatures and low specific humidities aloft.

Arctic air is produced over the ice- and snow-covered parts of the arctic region during the colder seasons, when the earth's surface is a marked cold source. In the North American sector its seat, for dynamic as well as thermal reasons, lies on the average near Baffin Island, being separated from the polar air south of it by the American arctic front (Fig. 3)

Fronts and frontogenesis. Because of the earth's rotation a surface of density discontinuity, a front, seeks a tilted position of dynamic equilibrium. A front is defined as a dynamically important, tilting layer of transition between two air masses of markedly different origin, temperature, density, and motion Colder air lies as a wedge below the warmer; therefore, the front will coincide with a trough or bend in the moving isobaric system, at whose passage the wind will veer, and as a rule with falling pressure ahead and stationary or rising pressure behind. Frictionally produced convergence, or static and inertial instability, or both, preferably within the warmer air, favors the ascent of air along the front. As a result a vast cloud mass, a cloud system, may form, with an area of continuous precipitation near the front (Fig 6).

The main fronts reach into the stratosphere (see Fig. 4) and have a horizontal extension of several thousand kilometers. They originate as links in the atmospheric circulation, within frontogenetic zones, between two stationary anticyclones (highs) and two cyclones (lows) when the axis of stretching, or confluence, of this deformation field (Fig. 5) has some extension west to east. This pattern is in its turn determined by the large-scale orography of the earth. Because of the north-south tempera u contrast, air masses of different temperatur

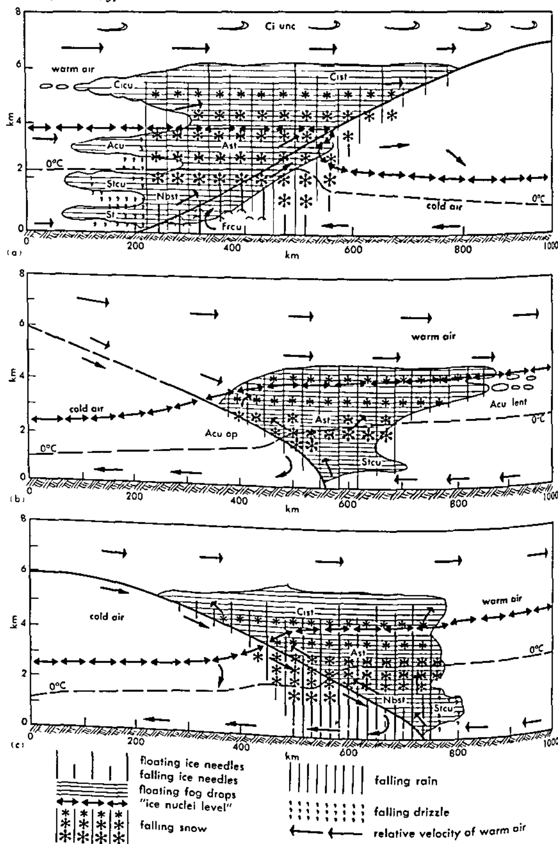


Fig. 6. Schematic cross sections of three kinds of fronts. (a) Warm front. (b) Fast-running cold front. (c) Slow-moving cold front. Acu, altocumulus; Acu lent, altocumulus lenticularis; Acu op, altocumulus opacus; Ast, altostratus; Ci unc, cirrus uncinus; Cicu, cirrocumulus; Cist, cirrostratus; Frcu, fractocumulus; Nbst, nimbostratus; St, stratus; Stcu, stratocumulus.

cific humidity, and density are then brought together in a front zone along this axis. On the warm side of the polar fronts, which together more or less encircle the hemisphere, the thermal wind, as discussed in dynamical meteorology below, corresponding to the mass distribution of the fronts, implies a narrow band of intense flow near the tropopause level, a jet stream. The thermal wind is most pronounced at middle and higher latitudes, where the front is marked, steep, and usually reaches into the stratosphere (Fig. 5). There, a main front and its jet together represent a zone of maximum potential and kinetic energy. Below, in lower latitudes, the frontal tilt is often small, and there is an outflow of polar air into the tropics.

A front may appear as warm or cold. At a warm front, the warmer air gains ground and slides evenly upward above the retreating cold-air wedge (tilt or slope 1:200 to 1:100), producing a wedge-shaped upglide cloud system (Fig. 6*a*). At the approach of a marked warm front, therefore, a typical cloud sequence invades the sky: cirrostratus → altostratus → nimbostratus, the last yielding continuous and prolonged precipitation ahead of the front line.

At a cold front the rather steep cold-air wedge (with a slope of 1:100 to 1:50) pushes forward under the warm air and forces it upward according to one of the two flow patterns shown in Fig. 6*b* and *c*. At the approach of a cold front of the more common type (Fig. 6*b*) there is, therefore, another typical cloud sequence: altocumulus, partly lenticular, rapidly thickening into nimbostratus. The precipitation is usually more intense but of shorter duration than with the warm front.

Waves and vortices (the weather). The actual atmospheric states affecting the weather may be regarded as composed of the general circulation and its disturbances. Together they form a multitude of weather mechanisms, some of which have already been described, in which the water-vapor cycle (evaporation → transport and lifting → condensation → precipitation) evidently has a fundamental role (*see also* CLOUDS; CLOUD PHYSICS). The atmospheric disturbances consist of a spectrum of waves of different wavelengths λ (Table 2), corresponding circulations (vortices), or both. Only the larger of these are studied synoptically. By their size the circulations may be classified as planetary (or geographical), secondary, or tertiary:

1. The long waves forming in the front ↔ jet zone, that is, the region of maximum energy, may represent a steady state (see section on dynamical meteorology) when their wavelength corresponds to about 4 circumpolar waves at low, 3 at medium, and possibly 2 at high latitudes. The planetary waves have $\lambda \sim 10{,}000$ km. Thereby they partly determine the shape of the general circulation (see Fig. 3). The shorter long waves ($3000 < \lambda < 8000$ km) propagate slowly eastward and, because of inertial instability, mostly develop as shown in Figs. 8 and 10 through the stages wave → tongue → cut-off vortex. Since the air aloft flows rapidly through these quasi-stationary long-wave patterns, isotherms will approach coincidence with the streamlines and isobars. Thus, equatorward tongues and cutoffs will be cold and will tend to coincide with lows (at least aloft); the poleward ones will tend to be warm and coincide with highs. The former will contain a polar-air hourglass (at least in higher latitudes) or dome, possibly with an arctic-air dome inside it (Fig. 9). Correspondingly, the poleward tongues and cutoff vortices will contain tropical air and a much higher tropopause.

2. The secondary waves are the short waves ($1000 < \lambda < 3000$ km). The short waves in a front ↔ jet system, when unstable, will form secondary circulations developing according to the scheme of Fig. 7, that is, initial front wave → young cyclone → initial occlusion → backbent occlusion. Important secondary circulations also form outside the front ↔ jet systems: the easterly wave, the tropical hurricane, and the convective system, occurring primarily in lower latitudes, the last two without an obvious preceding wave stage.

Table 2. Atmospheric waves

Name	Type*	Wavelength, km
Ultrasound	C	$<2 \times 10^{-5}$
Ordinary sound (tones)	C	2×10^{-5} to 10^{-2}
Explosion waves	C	10^{-2} to 10^{-1}
Helmholtz waves (Sc und, Ac und)	G	10^{-1} to 1
Short lee-waves (Ac lent, Cc lent)	G	1 to 2×10
Long lee-waves (nacreous clouds, precip.)	G(I)	2×10 to 10^2
Short jet-waves (frontal)	GI(V)	5×10^2 to 5×10^3
Long jet-waves	V	5×10^3 to 10^4
Tidal waves	(G)	$\leqq 2 \times 10^4$

* C, compression, longitudinal, G, gravitational, I, inertial, and V, vorticity-gradient, the last three are transversal

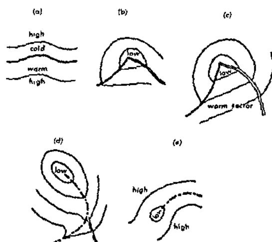


Fig. 7. (a-e) Life history of extratropical cyclones, Northern Hemisphere

3. The tertiary circulations, barely observable by the ordinary synoptic network, require a meso-scale network ($d < 10$ km), time sections, or both for a detailed study and forecast. Weather mechanisms of this size are land and sea breezes, mountain and valley winds, katabatic and other local winds (foehn, chinook, bora), local showers, tornadoes, orographic cloud and precipitation systems, and lee waves. Moreover, numerous different orographic factors and the daily period of radiation affect most meteorological elements and weather mechanisms. Their detailed study is therefore basic for both climatology and weather forecasting.

Life cycle of extratropical cyclones. An extratropical cyclogenesis starts as a wavelike front bulge. Ahead of the wave the front becomes a warm front, behind it a cold front. These two sections then flank a warm tongue (see Fig. 7*b*) moving along the front. At first there is only a shallow low around the tip of this warm sector, but even the passage of such an initial frontal wave may cause sudden, severe, and unexpected weather changes. If the front wave has an appropriate size ($\lambda \sim 1500$ km, amplitude $\geqq 200$ km), it will usually be unstable, narrow gradually, and at last overshoot as does a breaking sea wave. The cold front overtakes the warm front at the ground, and the warm-sector air is lifted and spreads aloft. As a result the common center of gravity of the system sinks, and potential energy is transformed into the kinetic energy of an increasing cyclonic circulation As long as this front occluding continues, the circulation increases, and the ensuing low deepens; whereas, when the warm sector has disappeared from the interior of the main cyclone, the latter will weaken, and the low fill; these features serve as good, physically comprehensible, prognostic rules. The occluded front is called an occlusion. The occluding process gives the clue to the life history of cyclones and anticyclones (Fig. 7). At its last phase the occluded front often lags behind and bends (Fig. 7*d* and *e*), a "false warm sector" forming between the direct and the back-bent occlusion. Several cyclones (and lows), together constituting a cyclone series, may form on one and the same front, moving with the upper, steering current, the first one being in the most advanced stage of development (see Fig. 8).

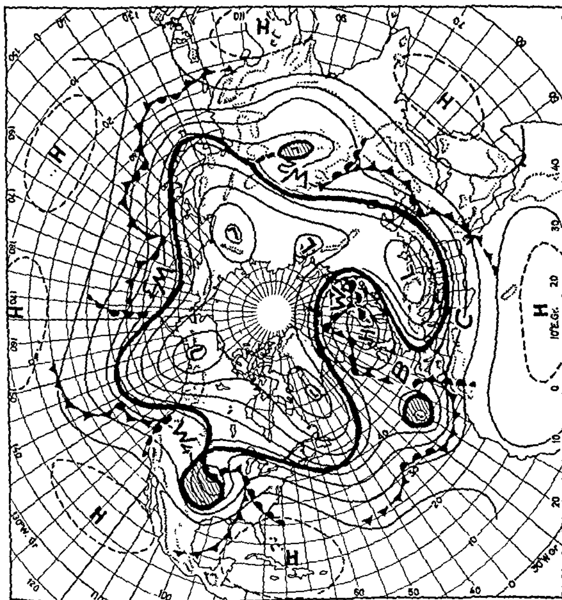
Important front ↔ jet systems in the Northern Hemisphere are the North Pacific polar front and the North Atlantic polar front (Fig. 3). The former extends in winter on the average from near the Philippines north of Hawaii to the northwest coast of the United States; the cyclone series forming in it then brings wet and unsteady weather to the western part of North America. The latter front, extending in winter on an average from near the Bermudas to England, plays the analogous role for almost all Europe. In summer both these systems are much weaker and lie on an average further north. The North American arctic front, and the European arctic front (Fig. 1), are fundamental for the weather in their vicinity. The former accounts for many severe weather vagaries, such as the blizzards and glaze storms of northeastern United States, and the most severe cold waves and their killing frosts further south; it is thus of special importance to North American weather forecasting. The production of real arctic air as defined here seems to end in summer, or at least in July. On the other hand, the corresponding air mass in the Southern Hemisphere, antarctic air, and the antarctic fronts, exist throughout the year, the antarctic ice plateau being an enormous cold source even in summer. Evidently most migrating extratropical precipitation and storm areas originate at front ↔ jet systems, separating air masses of radically different motion and weather type. Therefore, these systems are of utmost importance to weather forecasting, being the real atmospheric zones of danger and main sources of the salient aperiodic weather changes of synoptic size outside the tropics.

Large-scale tropical weather systems. In the tropics (and in the subtropics in summer) nonfrontal convective and other weather phenomena may grow to hundreds of kilometers in width; consequently, they can be studied and forecast individually by synoptic methods. Both on land and at sea, short waves ($\lambda \sim 500$ km) form in the frontless trade winds outside the doldrums. These easterly waves propagate slowly westward, showing a characteristic intensification of the shower activity in their eastern part.

Over tropical seas in late summer and early fall (Northern Hemisphere, August–October; Southern Hemisphere, February–April), conditions exist that favor the formation of tropical hurricanes. Tropical hurricanes are cyclonic vortices (smaller but much more intense than the extratropical ones) of 100–400-km width; they have wind velocities often exceeding 50 m/sec (100 knots), 100–500 mm total rainfall, and low central pressure. They move on the whole poleward, often recurving around a subtropic high. Necessary conditions for their formation seem to be (1) air (and sea-surface) temperature above, say, +27°C, implying enough labile energy to drive such large-scale convections, (2) a preexisting cyclonic motion and frictional inflow (either at the tropical front or in an easterly wave), needed to order, and possibly trigger, the convections, (3) a divergence mechanism aloft to dispose of the air that converges and rises in their interior, possibly also triggering their formation, (4) sufficient Coriolis force (that is, the hurricanes cannot be too near the Equator), and (5) no disturbing land surface within the area of formation. Whether these conditions are sufficient is not known. The widespread destructive power of hurricanes (shore flooding, wind pressure, downpour) makes their study a major task of tropical synoptic meteorology and weather forecasting. Tracking those already formed nowadays can be done with radar and reconnaissance flights, but discovering

their imminent formation is still an unsolved problem. *See* HURRICANE.

Over land, conditions 1 and 2 never lead to hurricane formation but may instead—within the tropics, and also in warmer seasons further poleward—favor the formation of a migrating convective system, with a pseudo cold front (in the United States also a forerunning squall line) at its outer edge; these systems have roughly the same extent and precipitative power as a tropical hurricane.



polar front aloft near mid-troposphere

polar front at earth's surface

pressure-line aloft

H L high and low aloft

ground > 2000 m above sea level

W warm ridge of upper long wave

C cold upper trough

B blocking

L Cut-off polar-air dome

Fig. 8. Schematic circumpolar upper pressure pattern in relation to the polar fronts at the earth's surface (After E. Palmen)

Outside the tropics they mainly form in the warmer seasons in cyclonic warm sectors, especially in the midwestern United States, where they provide the main water supply. Because of the flood hazards, soil erosion, and other aspects, and the sudden violent squalls, or even tornadoes, sometimes attending the pseudo cold front, these systems, therefore, form another major problem of weather forecasting. Within the tropics, migrating convective systems, like the easterly waves, will cause more variation of weather from day to day than is generally recognized, whereas the small-scale showers usually have a daily period; together they constitute the equatorial rains.

Apart from the polar-front jet (meandering between 30° and 70° lat), there is a subtropical jet, and a corresponding front, at about 30° lat near the tropopause level (~15 km altitude). As a rule it makes only small meridional excursions. Therefore, it shows up (instead of the polar-front jet) as the main wind maximum aloft (at 30° lat) in the average meridional cross section of Fig. 2. But for the same reason, and because its front is confined to the tropopause region, its influence on daily weather events, at least in the Tropics, seems to be slow and diffuse.

Circumpolar aerology after 1940. Since 1940, the technical facilities of ... nder- ... o the ... 1960 there was a network of about 400 aerologic stations around the Northern Hemisphere; twice a day these stations observe and signal the pressure, temperature, humidity, and wind in the atmosphere at least up to the 100-millibar level (15 km altitude). The vast gaps formerly in the meteorological network over the oceans are now partly bridged by stationary weather ships, or ocean weather stations (nine of them in the North Atlantic), with complete air-sounding equipment. This huge technical improvement has been followed by equally outstanding scientific achievements in the understanding of the dynamics and thermodynamics of the free atmosphere. The weather mechanisms and weather processes described above, which before 1940 had been mainly observed and studied within the lower half of the troposphere, can now be treated in their entirety and fitted into their general relationships with the rest of the atmosphere. On the whole, an intimate interaction takes place between the disturbances seen on the ordinary synoptic map (surface map) and the upper large-scale patterns shown by Fig. 8 and described above. Figure 9 shows this interdependance three dimensionally and in detail for disturbances of front ↔ jet systems (including the arctic front). Particularly, it shows the eight weather regions that are conditioned by such systems, these, naturally, being of fundamental importance to synoptic meteorology and weather forecasting.

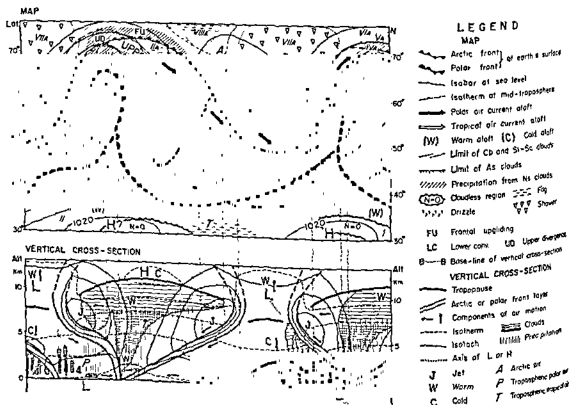


Fig. 9. Schematic three-dimensional model of middle-latitude front ↔ jet disturbances and their weather regions (fall and spring).

In winter, the polar-front jet lies on an average between 25 and 45° lat, in summer between 40 and 60° lat, at 7–11 km alt; it is only a few hundred kilometers broad, with wind speed surpassing 100 m/sec (200 knots). It displays two fundamentally different patterns: the zonal or high index type, where the jet has waves of only small amplitudes; and the meridional or low index type, where its meridional excursions are huge and often are cut off. Figure 10 gives an example of the isothermal, isobaric, and flow patterns at the midtroposphere, represented by the isotherms and isohypses ("contour lines") of the 500-millibar surface. It shows three long waves w_1 w_2 w_3, one deep trough T, and 4 cutoff lows L_1 L_2 L_3 L_4. Where the isotherms diverge markedly from the parallelism with the streamlines and "contours," there will be a considerable horizontal advection of colder and denser air or of warmer air (Fig. 10), and ensuing advective changes of pressure and wind at most levels. These observations gave another powerful tool for quantitatively forecasting the thermal changes of flow patterns. Calculating the vertical motion by the divergence and vorticity equations has now proved even better for this and similar purposes (as discussed below in dynamical meteorology).

Since the planetary waves represent a rather stable steady state, they offer no real forecast problem The shorter long waves will move east, conserving most of their absolute vorticity, and may therefore be forecast numerically, using a very simplified general model of the atmosphere, the barotropic model, where production and annihilation of kinetic energy is excluded. The propagation and development of the short front ↔ jet waves, that is the weather mechanisms, will to a certain extent be steered by this long-wave pattern. However, in these mechanisms frontal potential energy is transformed into cyclonic kinetic energy, and the short jet waves will often react upon the large-scale pattern. Such processes, which may lead to intense frontal cyclogenesis and the formation of large-scale cold tongues and cutoffs, confront numerical forecasting with considerable difficulties. However, the fact that the upper steering flow patterns are so simple and move much more slowly than the lower ones has proved a great help to short-range forecasting and has made an extended forecasting (2–5 days) possible on a synoptic-physical basis Today, attempts are made, with some success, to use baroclinic models of the atmosphere as well, and to incorporate the effect of heat sources and large-scale orography into the numerical short-range and extended forecasting. *A main obstacle to real long-range forecasting* (7–30 days) is the fact that the causes for

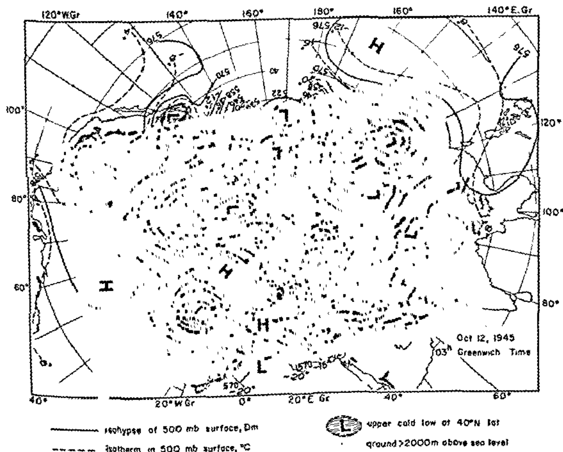


Fig. 10. Circumpolar pressure, flow and temperature pattern aloft, Oct. 12, 1945, at 03h z.

a definitive change from the zonal to the meridional upper flow pattern are not yet known.

Because of radar, continuous detailed mapping and tracking of rain areas is gradually being introduced. Analogously, a photographic and television cloud survey from satellites, already tried experimentally in the United States, promises to become of great help for studying the large-scale development of weather regions. [T.BE.]

DYNAMICAL METEOROLOGY

This branch of meteorology is the science of naturally produced motions in the atmosphere and of the related distributions of pressure, density, temperature, and humidity. It forms the main scientific basis of weather forecasting and climatology—*see* WEATHER (FORECASTING AND PREDICTION).

The motions in the atmosphere are caused by solar heat. The combined process of absorption and emission of radiation causes a distribution of heat and cold sources which continually disturbs the equilibrium and maintains the motion. During this process, the volume and pressure of individual air particles change with time so that expansion takes place, on the average, at higher pressure than contraction. As a result, heat is converted into mechanical energy, just as in a man-made heat engine. The bulk of the mechanical energy thus released is used to overcome friction within the atmosphere itself; only a very small fraction is spent to do work on the ocean surfaces, maintaining ocean currents and waves. The rotation of the earth exerts a profound influence on the motions in the atmosphere. Atmosphere-like motions, resembling major circulations of the air, have been reproduced in the laboratory in rotating tanks of water exposed to [illegible]

[illegible] pressure, air density, and three components of air velocity (latitudinal, longitudinal, and vertical in terrestrial terms). These basic quantities are regarded as functions of three space coordinates and time, and as such they characterize a sequence of states, or a motion process in the atmosphere. It is sometimes necessary to add humidity as a sixth variable.

The central problem of dynamical meteorology is the prediction problem: from a known initial distribution of the basic variables in space to determine the subsequent states. A related problem concerns the existence of possible steady states and their stability properties. A problem relevant to climatology is that of determining the statistical characteristics of motions over long periods of time.

These problem[illegible] [illegible] equations of motion, expressing the principles of Newtonian dynamics; the equation of continuity, expressing conservation of mass; and the thermodynamic energy equation, expressing the first law of thermodynamics. A sixth equation, expressing the conservation of the mass of water substance, must be added if humidity is taken into consideration. In addition, the variables must satisfy certain boundary conditions at the earth's surface and the upper limit of the atmosphere. These equations will under certain conditions form a closed system which, in principle at least, determines the solutions to the problems mentioned above.

A basic difficulty in the theory results from the fact that the equations are nonlinear and, therefore, not tractable by standard mathematical methods; two solutions cannot be added to form a new solution. The complexity of the motions presents another difficulty. Superimposed upon the large-scale motion systems revealed by weather maps is a fine structure of motions of all scales down to tiny eddies of millimeter size. Such motion fields can not be handled; it is necessary to deal instead with generalized or smoothed fields, where details smaller than a certain scale have been ironed out. As a consequence of the nonlinearity of the equations, these details will still exert a certain influence upon the larger-scale motions. This influence can only be taken into account, in a statistical sense, in the form of so-called eddy terms, which express the transport of momentum and energy carried out by small-scale eddies.

Hydrodynamic equations. The form of the equations of motion is complicated by the fact that the earth is a sphere. However, the principal features of the dynamics of air motions will still emerge if for simplicity the earth is considered as flat. In this case a right-handed cartesian coordinate system (x,y,z) may be used, with the z axis pointing towards zenith. The equations of motion then are

$$\frac{Du}{Dt} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + fv + \frac{1}{\rho}\frac{\partial \tau_x}{\partial z} \quad (1)$$

$$\frac{Dv}{Dt} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - fu + \frac{1}{\rho}\frac{\partial \tau_y}{\partial z} \quad (2)$$

$$\frac{Dw}{Dt} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g \quad (3)$$

where

$$f = 2\Omega \sin \varphi \quad (4)$$

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z} \quad (5)$$

The notation is as follows: t denotes time; u, v, w, velocity components in directions of x, y, z; p, pressure; ρ, density; g, acceleration of gravity; f, Coriolis parameter; $\Omega = 0.7292 \times 10^{-4}$ rad/sec angular velocity of the earth; φ, latitude; τ_x, τ_y horizontal eddy stresses in the x and y direction on a horizontal surface due to vertical eddy transport of horizontal momentum; D/Dt individual derivative, or rate of change as experienced by a moving air particle.

Equations (1) and (2) state that the acceleration in the direction of x and y, respectively, i[illegible]

equal to the sum of the pressure force, the Coriolis force (or deflecting force of the earth's rotation) and the force of eddy friction, per unit mass. Equation (3) states that the vertical acceleration equals the sum of the vertical pressure force and the force of gravity, per unit mass. Certain terms of little dynamic significance (Coriolis terms involving $\Omega \cos \varphi$, certain small eddy terms, and molecular viscous forces) have been omitted.

The equation of continuity

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho u)}{\partial x} - \frac{\partial(\rho v)}{\partial y} - \frac{\partial(\rho w)}{\partial z} \qquad (6)$$

states that the increase of the mass contained within any volume must equal the inflow of mass across its boundary surface. *See also* FLUID-FLOW PRINCIPLES.

Thermodynamic equations. Air contains variable amounts of water vapor (up to a few per cent), and sometimes also suspended water droplets, ice particles, or both. It may be assumed that no condensation products are present as long as the partial pressure of water vapor is less than the saturation pressure with respect to a plane water surface. Such nonsaturated air behaves approximately as a single perfect gas. It satisfies the gas equation

$$p = RT\rho \qquad (7)$$

by means of which the absolute temperature T can be expressed in terms of the basic variables p and ρ; R is the gas constant referred to unit mass. The thermodynamic energy equation has for nonsaturated air the form

$$c_p \frac{DT}{Dt} - \frac{1}{\rho}\frac{Dp}{Dt} = q \qquad (8)$$

where c_p denotes specific heat at constant pressure, and q amount of heat absorbed (or lost, if negative) per unit mass of air per unit time, including the generation of heat by frictional dissipation of mechanical energy. The values of R and c_p depend slightly upon humidity, but usually the values for dry air may be used with sufficient accuracy; these are $R = 287\ \text{m}^2\ \text{sec}^{-2}\ {}^\circ\text{K}^{-1}$, and $c_p = 1004\ \text{m}^2\ \text{sec}^{-2}\ {}^\circ\text{K}^{-1}$.

A sample of nonsaturated air may become saturated by removal of heat or by a decrease in pressure. If this process continues, supersaturation is prevented by condensation of a proper amount of vapor into cloud droplets (the distinction between water and ice is of little importance for the dynamics), and the air remains saturated as long as water droplets are present. For such saturated air, a term representing the released heat of condensation must be added in Eq. (8). The effect of cloud droplets upon air density is usually negligible.

In the absence of heat sources ($q = 0$), Eqs. (7) and (8) give, in the case of nonsaturated air,

$$\frac{DT}{Dp} = \frac{RT}{c_p p} \qquad (9)$$

that is, the quantity

$$\theta = T\left(\frac{p_0}{p}\right)^{R/c_p} \quad (p_0 = 1000\ \text{millibars}) \qquad (10)$$

called the potential temperature ...

gram (or a similar diagram). In saturated air containing water droplets, compression will be accompanied by evaporation and the increase in temperature will be less. Changes of state of saturated air without heat sources are called saturated-adiabatic and may be represented by another family of curves in a pT diagram. *See also* THERMODYNAMIC PRINCIPLES; THERMODYNAMIC PROCESSES.

Motions without friction and heat sources. Many motion phenomena of short duration (up to a few days) are relatively unaffected by friction and heat sources and can be studied under the assumption that these effects are absent. In this case the equations permit a particularly simple solution representing a steady, straight, horizontal current. With the x axis chosen in the direction of the current, so that $v = w = 0$, and with u, p, and ρ independent of x and t and dependent on y and z only, Eqs. (1), (6), and (8) are identically satisfied, whereas (2) and (3) require

$$-\frac{1}{\rho}\frac{\partial p}{\partial y} - fu = 0 \qquad (11)$$

$$-\frac{1}{\rho}\frac{\partial p}{\partial z} - g = 0 \qquad (12)$$

Equation (11) states that the horizontal pressure force is balanced by the Coriolis force. The wind necessary to establish such a balance is called the geostrophic wind. At the equator the Coriolis parameter f vanishes, and geostrophic balance cannot be established. Equation (12), called the hydrostatic equation, states that the vertical pressure force is balanced by the force of gravity.

The density field is said to be barotropic if the density (and hence also T and θ) is uniform in each isobaric (constant pressure) surface; in the opposite case it is said to be baroclinic. It can be inferred from Eqs. (11) and (12) that baroclinicity is associated with a change of geostrophic wind with height, whereas in a barotropic current the wind remains the same at all levels (but may change with y). In particular, the state of permanent rest must be barotropic with horizontal isobaric surfaces.

A discontinuity surface, where u and ρ change abruptly, may exist in a steady current. It will be held in a slanting position towards the y axis, against gravity, by the action of differential Coriolis forces on its two sides. Pressure is continuous, but isobaric surfaces are refracted upward in passing through the surface. Atmospheric fronts may be considered as discontinuity surfaces of this kind.

Stability and wave motions. A steady state is ... the state le if a sig- e can de- ver small. velop ...

For a state which deviates only slightly from a known steady state, the basic nonlinear Eqs. (1) to (3), (6), and (8) turn into a set of linear perturbation equations in the small perturbation quantities which define the deviations of the variables u, v, w, p, and ρ from their steady-state values. The perturbation equations have particular solutions representing various kinds of wave motions, whose amplitudes may either remain small or grow indefinitely. Existence or nonexistence of amplifying ... to in- kinds

... ere in steady state, a particle is displaced vertically, its buoyancy and weight will as a rule no longer balance. Their resultant may pull the particle towards or away from its original position. In the former case, the state is said to be statically stable; this occurs when the lapse rate of temperature $(-\partial T/\partial z)$ in the undisturbed state is less than the adiabatic rate $(-DT/Dz)$ at which temperature changes for the displaced particle. In the opposite case, $(-\partial T/\partial z > -DT/Dz)$, the state is statically unstable. The adiabatic lapse rate $(-DT/Dz)$ equals 0.01C°/m for a nonsaturated particle, and somewhat less, depending on p and T, for a saturated particle. Static instability may result from heating by the underlying surface, and results in convection currents, which, in case saturation occurs in the rising currents, become visible as cumuliform clouds.

Inertial stability. In a steady current this condition results from excessive Coriolis forces acting on a chain of particles parallel to the current which have been displaced horizontally normal to the current. The criterion of stability is $f(f - \partial u/\partial y) > 0$. If the chain of particles is given a vertical as well as a horizontal displacement, static and inertial stability both become operative. Instability may still result for some directions of displacement, even when the current is statically as well as inertially stable, when the density field is sufficiently baroclinic.

In a statically and inertially stable atmosphere, there is a spectrum of possible ...

... s

parameter f. Theoretical study of other types of wave motions in a straight current have revealed other types of instability. Thus a certain type of wave of wavelength of the order 3000 km is found to amplify if the current is sufficiently baroclinic. This kind of instability is called baroclinic instability. The growing waves receive their energy from the potential and internal energy of the air; their structure is in many respects similar to the observed structure of intensifying cyclones. Another type of amplifying wave, which feeds on the kinetic energy of the current, may exist if the velocity profile of the current is such that $(-\partial u/\partial y)$ has a strong maximum at a certain value of y. This type of instability is sometimes called barotropic instability, since it may occur also in a barotropic current.

Cyclones in middle latitudes are observed to form in the polar front zone, characterized by pronounced baroclinicity and also by a strong maximum of the horizontal shear of westerly winds. Baroclinic and barotropic instability may both operate to cause the sudden growth of cyclones that is often observed. See BAROCLINIC FIELD; BAROTROPIC FIELD.

Theory of large-scale motions. For motion systems of horizontal scale of the order 1000 km or more, the accelerations are so weak that the equilibrium conditions (11) and (12) are approximately valid; these motions are thus quasi-static and quasi-geostrophic. Equation (12) holds with great accuracy and may replace (3); Eq. (11) is usually valid within 10–20% error.

The vertical vorticity $\zeta = (\partial v/\partial x) - (\partial u/\partial y)$ represents twice the speed of rotation of an air particle with respect to a vertical axis (reckoned positive when counter-clockwise). The horizontal divergence $\delta = (\partial u/\partial x) + (\partial v/\partial y)$ measures the relative rate of expansion of an infinitesimal horizontal area moving with the air. These two quantities may always be used instead of u and v to characterize the horizontal velocity distribution. It is significant that the large-scale velocity fields are approximately characterized by ζ alone, δ giving only a small modification. The evolution of the velocity field is therefore approximately determined by the vorticity equation, which can be derived from Eqs. (1) and (2). Somewhat simplified, it reads

$$D(f + \zeta)/Dt = -(f + \zeta)\delta \tag{13}$$

f may be considered as the absolute vertical vorticity (in a nonrotating frame) due to the rotation of the earth; and $(f + \zeta)$ is therefore the total absolute vorticity. Equation (13) expresses the familiar mechanical principle that the absolute rotation of an air particle around a vertical axis will speed up when the particle contracts horizontally. For large-scale motion the continuity equation (6) may be written approximately as

$$\partial(\rho^* w)/\partial z = -\rho^*\delta \tag{14}$$

where ρ^* denotes a standard density distribution depending on z only. Equation (14) states that horizontal contraction is associated with vertical stretching of air columns, and horizontal expansion with vertical shrinkage.

In a current moving up a mountain slope, air columns must shrink vertically, and hence, expand

horizontally; the absolute vorticity must therefore decrease. The opposite process takes place when the motion is downslope. This explains the observed predominance of anticyclonic vorticity ($\zeta < 0$ in the northern hemisphere) over large mountain ranges which are crossed by air currents.

Over level country, there is no such net vertical stretching or shrinking through the whole atmosphere, and there is therefore a tendency towards conservation of absolute vorticity, that is

$$f + \zeta = \text{const} \qquad (15)$$

for each particle. Such conservation is not strictly realized, notably because vertical stretching may occur at some levels, and shrinkage at others; but Eq. (15) can still be used as a crude approximation at levels of about 5 km above sea level. According to (15), ζ must decrease for particles moving northward toward higher latitudes and higher values of f, and vice versa. A consequence of this is that the large-scale motion systems have a tendency to be propagated westward relative to the air. The effect can be studied even on a flat earth, in a cartesian coordinate system where f increases with y (northward) at a rate $\beta = df/dy$. A wave of wavelength L, superimposed upon a uniform westerly current of speed u, will then, according to Eq. (15), be propagated eastward at a speed c satisfying the Rossby formula

$$c = u - \frac{\beta}{4\pi^2} L^2 \qquad (16)$$

The speed of propagation thus depends upon wavelength; shorter waves move eastward slower than the air, and sufficiently long waves move westward. The waves are stationary for a wavelength $L = 2\pi\sqrt{u/\beta}$, a quantity that increases with the current speed u and with latitude. These relations are in fair agreement with the observed general behavior of such motion systems.

A more accurate treatment of large-scale motions is based on the hypothesis that the vertical motions will be so adjusted that the approximate geostrophic balance is maintained everywhere. This requirement suffices to determine the vertical motion, and hence δ, when the distribution of ζ is known. By entering this value of δ into the vorticity Eq. (13), a more accurate integration of this equation is possible. The method permits the incorporation of the effects of friction and heat sources. However, to save tremendous labor and time, the integrations must be carried out numerically by means of electronic computers. [A.E.]

Bibliography: H. R. Byers, *General Meteorology*, 3d ed., 1959; A. Eliassen and E. Kleinschmidt, Dynamic Meteorology, in S. Fluegge (ed.), *Handbuch der Physik*, vol. 48, 1957; C. L. Godske, T. Bergeron, J. Bjerknes and R. C. Bundgaard, *Dynamic Meteorology and Weather Forecasting*, 1957; T. F. Malone (ed.), *Compendium of Meteorology*, 1951.

Meter (unit)

A unit of length used mainly in scientific work. The standard meter is the international standard unit of length, defined as the distance between two fine lines engraved on the international prototype meter bar when the bar is at the temperature of melting ice. The international prototype is a platinum-iridium bar kept at the International Bureau of Weights and Measures in Sèvres, France. Because of the possibility of damage or destruction of the prototype, it has been suggested that a supplementary definition of the meter be given in terms of the wavelength of light corresponding to the red line of cadmium or to the green line in the spectrum of the mercury isotope 198. Thus,

1 m = 1,553,164.13 cadmium wavelengths
= 1,831,249.21 mercury-198 wavelengths

where the wavelengths specified are in air at 760 mm pressure and 0°C. *See* WAVELENGTH STANDARDS. [D.WI.]

Methane

A member of the alkane or paraffin series of hydrocarbons, formula, CH_4. It has been called marsh gas because it is formed by anaerobic bacterial decomposition of vegetable matter in swampy land. To coal miners, it is known as "fire damp" because mixtures with air are explosive. It is a major constituent of natural gas (50–90%) and of coal gas. It is formed in large quantities by the activated sludge process of sewage disposal. It freezes at −182.6°C and boils at −161.6°C.

Besides its use as a fuel, methane is important as a source of organic chemicals and of hydrogen. Its reaction with steam at high temperatures in the presence of catalysts yields carbon monoxide and hydrogen (synthesis gas) which can be catalytically converted to liquid alkanes (Fischer-Tropsch process) or to methanol and other alcohols. The catalytic reaction of the synthesis gas with olefins yields higher alcohols (the Oxo process), and reaction with steam produces additional hydrogen and carbon dioxide.

The incomplete combustion of methane with air produces finely divided carbon, called carbon black, hundreds of millions of pounds of which are used annually as a reinforcing and filling agent in compounding rubber and as a pigment in black printing ink.

Chlorination of methane yields methyl chloride, methylene chloride, chloroform, and carbon tetrachloride. *See* ALKANE; FISCHER-TROPSCH PROCESS; HYDROFORMYLATION. [L.S.]

Methanol

An alcohol, CH_3OH, that is also known as wood alcohol, methyl alcohol, and carbinol. It was first isolated in 1661 by Robert Boyle from the products of the destructive distillation of wood. From that day to the present, methanol has been a leadin

dustrial organic chemical. In 1956, 1,600,000,000 lb were produced in the United States alone.

This simplest of the alcohols is a colorless, mobile liquid of pungent odor and taste and has the following physical properties: molecular weight, 32.04; boiling point, 64.8°C; melting point, −97.8°C; specific gravity, 0.7924. It is completely miscible with water and most organic liquids, and it is highly toxic and flammable.

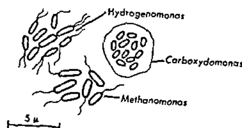
Manufacture of methanol today is almost exclusively by the high-pressure (100–600 atm) reduction of mixtures of carbon monoxide or carbon dioxide, or both, with hydrogen over metal oxide catalysts at 250–400°C. Lesser amounts are produced by partial oxidation of hydrocarbons from natural gas and from pyroligneous acid, a product of the destructive distillation of wood.

The chemistry of methanol is in general typical of all primary aliphatic alcohols. Its major use is as an intermediate in the synthesis of formaldehyde, which is used in the manufacture of plastics and numerous organic derivatives. Methanol is readily converted to methyl acetate, methyl chloride, and the methyl amines, all of which are important industrial organic chemicals. Methanol itself is used as a solvent, extractant, and special fuel, as well as a denaturant for ethyl alcohol and a component of antifreeze mixtures. *See* ALCOHOL. [J.W.L.]

Methanomonadaceae

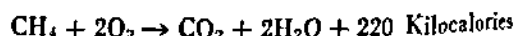
A family of bacteria of the suborder Pseudomonadineae which can live by deriving energy from the oxidation of methane, hydrogen, or carbon monoxide. The bacteria are gram-negative, rod-shaped, and motile with polar flagella. The species are not photosynthetic. The family includes three genera: *Methanomonas*, *Hydrogenomonas*, and *Carboxydomonas*. In 1956 M. T. Dworkin and J. W. Foster and in 1958 E. R. Leadbetter and J. W. Foster proposed to abandon the genus *Methanomonas*, and to incorporate it in the genus *Pseudomonas*; the ability to oxidize methane does not appear to be a sufficient reason for the creation of a separate genus. If this be accepted, one should, on the same basis, merge the other two genera with *Pseudomonas*.

A consequence of such a procedure would be the elimination of the entire family of Methanomonadaceae; to a number of bacterial systematists this would seem a rational step.



Genera of the Methanomonadaceae. (V. B. D. Skerman)

Methanomonas. A genus of bacteria which can oxidize methane. There is only one known species, *Methanomonas methanica* (Söhngen) Orla-Jensen, which was described by the Dutch bacteriologist N. L. Söhngen in 1906. He isolated it by inoculating a liquid mineral medium with soil or ditch water and incubating under a mixture of air and methane. The organism is a short rod, 2–3 µ by 1.5–2 µ, and cells in very young cultures are motile with one polar flagellum. In old cultures the cells shorten and become coccoid. It will grow in liquid media containing no organic matter, if methane and oxygen are present in the atmosphere above the liquid. Under these conditions it uses the reaction



$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O + 220 \text{ Kilocalories}$$

as a source of energy, and methane as its source of carbon for growth. [J.M.]

Hydrogenomonas. The species are gram-negative, polarly flagellated if motile, and widely distributed in nature. They differ from those of the related genus *Pseudomonas* in the ability to grow exclusively with carbon dioxide as a carbon source, utilizing energy from the reaction: $2H_2 + O_2 \rightarrow 2H_2O$. Species are differentiated as aerobic, requiring oxygen and growing best at high oxygen tension such as 20% or more, or microaerophilic, needing oxygen but growing best at or even requiring reduced oxygen tensions such as 10% or less. Limited investigation suggests a mode of carbon dioxide fixation analogous to that of the green algae. These bacteria are facultative autotrophs (can use either carbon dioxide or organic substances as carbon sources); most species grow well with organic compounds. However, inability of a culture to revert to autotrophy after such heterotrophic (grows at the expense of organic compounds as carbon source) cultivation is frequently observed. The ability to grow with hydrogen, oxygen, and carbon dioxide is not limited to *Hydrogenomonas* but is also found in species of *Mycobacterium*, *Bacillus*, *Micrococcus*, *Pseudomonas*, *Streptomyces*, and the algae.

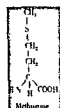
Carboxydomonas. The genus *Carboxydomonas* contains one incompletely described species, *C. oligocarbophilus*. The organism is an obligate autotroph (requiring inorganic compounds as nutrients) and is able to oxidize carbon monoxide to carbon dioxide as the source of energy for growth. K. Lantzsch described an actinomycete, which be regarded as identical with *C. oligocarbophilus*, that effects the same reaction. A. Kistner questioned the original techniques and observations, reinvestigated the problem and verified carbon monoxide oxidation with a newly isolated organism. Kistner's organism can also grow as a hydrogen bacterium; consequently he named it *Hydrogenomonas carboxydovorans*. Possible connections between *Carboxydomonas oligocarbophilus*, Lantzsch's acti-

nomycete, and *Hydrogenomonas carboxydovorans* are at present not clear and a critical study of the carbon monoxide-oxidizing bacteria is required. *See* BACTERIA, TAXONOMY OF; PSEUDOMONADINEAE.

[C.R.B.]

Bibliography: M. Dworkin and J. W. Foster, *J. Bacteriol.*, 72:646, 1956; E. R. Leadbetter and J. W. Foster, *Arch. Mikrobiol.*, 30:91, 1958.

Methionine



Physical constants of the L isomer at 25 C
pK_1 (COOH) 2.28 pK_2 (NH_3^+) 9.21
Isoelectric point 5.74
Optical rotation $[\alpha]_D(H_2O)$ −10.0 $[\alpha]_D$(5 N HCl) +23.2
Solubility (g/100 ml H_2O) 3.35 (DL)

An amino acid considered essential for normal growth of animals. The amino acids are characterized physically by the following: (1) the pK_1, or the dissociation constant of the various titratable groups; (2) the isoelectric point, or pH at which a dipolar ion does not migrate in an electric field; (3) the optical rotation, or the rotation imparted to a beam of plane-polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml; (4) solubility. *See* EQUILIBRIUM, IONIC; ISOELECTRIC POINT; OPTICAL ACTIVITY; SPECTROPHOTOMETRIC ANALYSIS.

A red color is formed when methionine is treated with nitroprusside in alkaline solution and then acidified. Methionine is an important methyl group donor in transmethylation reactions. For this purpose, it must first be activated by adenosinetriphosphate (ATP), forming *S*-adenosylmethionine. This compound readily transfers its methyl group to suitable acceptors, leaving *S*-adenosylhomocysteine. Active methyl groups appear to originate during methionine biosynthesis, because there is evidence that homocysteine receives a hydroxymethyl group from serine, following which reduction to the methyl group of methionine occurs (*see* AMINO ACIDS).

Several metabolic degradation pathways are known:

1. Homoserine may be formed by reversal of the biosynthetic reactions described for microorganisms. Homoserine then can be degraded further by two alternative routes: (*a*) oxidative deamination to α-keto-γ-hydroxybutyrate, followed by cleavage to pyruvate plus formate; (*b*) nonoxidative deamination to α-ketobutyrate.

2. Homocysteine, arising by demethylation, may be oxidized to homocysteic acid or desulfhydrated to α-ketobutyrate, hydrogen sulfide, and ammonia.

3. Methionine may be deaminated oxidatively to α-keto-γ-methiolbutyric acid, which is further degraded to α-ketobutyrate, methylmercaptan, and ammonia.

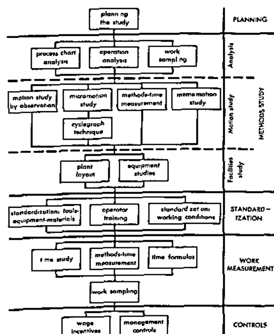
[E.A.AD.]

Methods engineering

A technique used by management to improve working methods and reduce labor costs in factories, offices, and wherever human effort is required. It analyzes each operation of a given piece of work to eliminate every unnecessary operation and to approach the quickest and best method of performing each necessary operation; it includes the standardization of equipment, methods, and working conditions; it trains the operator to follow the standard method. After all this has been done, it determines by measurement the number of standard hours in which an operator working with standard performance can do the job. Finally, it usually, although not necessarily, devises a plan for compensating labor that encourages the operator to attain or to surpass standard performance (*see* WAGE INCENTIVES).

Methods engineering finds the best way to do a job and then motivates men to use it. The illustration shows the techniques of methods study. Where alternate techniques are shown, the choice of the one to be used on a study is influenced by the repetitiveness, labor content, labor cost, and estimated life of the job. Each study must be planned at the outset so that time and money spent making it are proportional to and always less than the savings which result.

Methods are improved by observing what is currently being done and then developing better ways of doing it (*see* METHODS STUDY). If a new job not yet in production is being studied, a practical method of doing it is first visualized. Then better



Graphic analysis of the elements of methods engineering.

ways of accomplishing the same objective are sought. The process is one of starting with an existing method, either actual or visualized, analyzing it so that it is clearly understood in detail, and then searching for better ways of doing parts or all of the work.

Where work has not previously been studied, methods engineers repeatedly find that much of what is being done is entirely unnecessary. Sometimes whole operations are eliminated merely by asking the question "Why is this being done?" and recognizing that there is no sound answer. Primary techniques used to discover such situations are process chart analysis, operation analysis, and work sampling.

When all unnecessary work has been eliminated and obvious methods improvements have been made, the methods engineer employs a secondary motion study to analyze individual motions if the repetitiveness of the job justifies it. Motion study usually leads to the devising of still easier and less fatiguing ways of doing the job. The quickest way of making a motion study is by direct observation. Where the repetitiveness of the job justifies its use, micromotion study is more effective. The methods engineer takes a motion picture of the job and analyzes the resulting film in detail. He then improves the method by identifying and correcting ineffective portions of the motion pattern.

The related cyclegraph technique is occasionally used for studying motion paths in greater detail. A three-dimensional pattern of a motion path is obtained by photographing a small electric light attached to the body member on which information is desired. The stereoscopic picture records for analysis the path of the light through space.

Systems of predetermined elemental times are useful for methods study (see WORK MEASUREMENT) Memomotion study uses a motion picture camera operating at slower than normal speeds. The technique is used for studying work where the operation cycle is irregular and of long duration.

Methods improvements are additionally sought from facilities studies including plant layout and equipment studies. Because of the knowledge of the work that the methods engineer gains from his studies, he is often able to improve layouts so as to reduce material handling time or to suggest improved equipment and processes.

Methods study is sometimes called work simplification. If the work is highly repetitive, it tends to lead to [illegible]

[illegible]

can be fo[illegible]
zation pr[illegible]
to do the work. Depending upon the nature and importance of the job, training can consist of verbal instructions, demonstrations at or away from the work place, written instruction cards, visual guidance with motion pictures, or a combination of these and still other training techniques. In some cases, attitude training, to develop "want to," accompanies training in "know how."

After the method has been made as efficient as is economically justified and after standardization has been accomplished, the job is ready for work measurement. When stop-watch time study is to be used, the methods engineer selects a suitable operator, explains the purpose of the study to him and makes his observations, usually rating the performance of the operator by judging the skill and effort he is exhibiting or by assessing the speed with which his motions are made as compared to a normal working pace. He makes the necessary calculations, adds allowances—usually determined from separate work sampling studies—for fatigue and personal, unavoidable, and special delays, and establishes a standard time for doing the work.

If a system of predetermined elemental times has been used during the methods study, a list of the motions required to perform the operation will have been developed. The methods engineer can then measure the work merely by assigning predetermined time values to each motion. When these are totaled and the necessary allowances are added, the standard time is obtained without need for further observation or measurement.

On nonrepetitive work, the cost of establishing standards by individual studies is usually prohibitive. In such cases, the methods engineer develops standard data or time formulas. With these compilations of detailed time data as table, chart, or algebraic formula, the methods engineer need only identify major variables affecting performance time such as size, weight, or the like and select appropriate data from the tables or substitute in the formula to establish a time standard. Although the cost of developing a time formula may be substantial, once developed, the job of establishing accurate and consistent standards becomes routine, often being accomplished in less than 5 min per standard. Time formulas are particularly useful in the measurement and control of indirect labor.

When time standards are available, they may be used for control purposes in connection with production planning and control, standard costs, budgetary control, top management controls, manning controls, machine loading, and operator performance control. They may further be used as the basis for incentive wage payment.

In summary, the procedures in the chart grouped under the category of methods study are used to find the best way to do a job. The procedures grouped under standardization, work measurement and controls are employed to get the operator to use the best method, once it has been found. See INDUSTRIAL ENGINEERING. [H B M]

Methods study

An industrial engineering procedure for finding better ways of doing work and of reducing costs. Methods studies have been made and substantial econo-

mies obtained in factories, offices, hotels, hospitals, retail stores, warehouses, utilities, and banks—in short, whenever human labor is needed to accomplish useful work.

Figure 1 shows the procedures used to study and improve a method. Most methods studies are made to reduce costs. Occasionally they are made to improve quality, to remove a production bottleneck, or to overcome shortages of skilled labor.

In making a methods study, the methods engineer studies the existing method, or, if it is a new job, the method he visualizes may be used. The analysis consists of taking the operation apart in ever smaller pieces. Each piece is studied in detail. When the engineer has learned all that he can about the operation, he discards unnecessary pieces and puts the remainder together in the best way he can. This procedure usually results in an improved method.

A fundamental of methods study is that with sufficient study, any method can be improved. It takes time and costs money to make a study; therefore, the methods engineer plans his studies so that he will spend enough time on each one to get as much as is practical of the potential savings but will not spend so much time that the cost of the study will be out of proportion to the savings. The most profitable jobs to study are those with the most repetition, the highest labor content (manual work as distinguished from mechanical or process work), the highest labor cost, or the longest life span measured in months or years.

Process chart analysis. A process chart is a graphic representation of events occurring during a series of actions or operations. It is useful for obtaining an initial understanding of a process. Process charts present a clear picture of a given process and often reveal unnecessary work or duplication of effort.

Events recorded on a process chart

	*Operation.* An operation occurs when an object is intentionally changed in any of its [illegible] storage. An operation also occurs when information is given or received or when planning or calculating takes place
	*Transportation.* A transportation occurs when an object is moved from one place to another, except when such movements are a part of the operation or are caused by the operator at the work station during an operation or an inspection
	*Inspection.* An inspection occurs when an object is examined for identification or is verified for quality or quantity in any of its characteristics
	*Delay.* A delay occurs to an object when conditions except those which intentionally change the physical or chemical characteristics of the object do not permit or require immediate performance of the next planned action
	*Storage.* A storage occurs when an object is kept and protected against unauthorized removal.

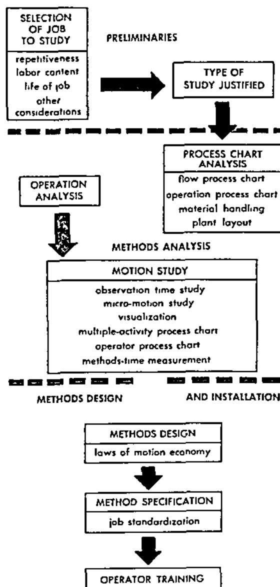


Fig. 1. Graphic analysis of the elements of methods study. (*From S. M. Lowry, H. B. Maynard, and G. J. Stegemerten, Time and Motion Study, 3d ed., McGraw-Hill, 1940*)

Process charts are of four major types: (1) operation process charts, (2) flow process charts, (3) multiple-activity process charts, and (4) operator process charts. The first two types are used for the analysis of processes involving a number of events or operations. The other two are used to analyze single operations in detail.

The events recognized in charting a process and the symbols for them are shown in the table.

An operation process chart represents the points at which materials are introduced into the process and the sequence of inspections and all operations except material handling. It is particularly useful in connection with plant layout studies. Figure 2 [illegible] operation process chart.

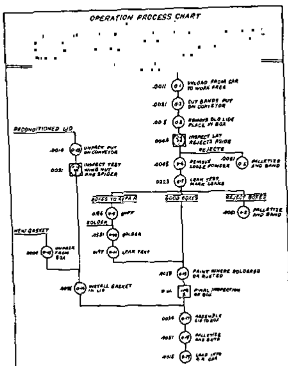
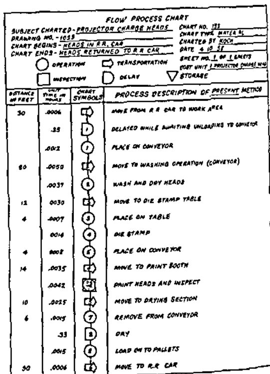


Fig. 2. Typical operation process chart.



FLOW PROCESS CHART

SUBJECT CHARTED - PROJECTOR CHARGE HEADS — CHART NO. 131
DRAWING NO. - 1023 — CHART TYPE MATERIAL
CHART BEGINS - HEADS IN R.R. CAR — CHARTED BY KOCH
CHART ENDS - HEADS RETURNED TO R.R. CAR — DATE 4 10 58
SHEET NO. 1 OF 1 SHEETS
COST UNIT 1 PROJECTOR CHARGE HEAD

OPERATION — TRANSPORTATION — INSPECTION — DELAY — STORAGE

DISTANCE IN FEET	UNIT TIME IN HOURS	CHART SYMBOLS	PROCESS DESCRIPTION OF PRESENT METHOD
30	.0006		MOVE FROM R.R. CAR TO WORK AREA
	.35	1	DELAYED WHILE AWAITING UNLOADING TO CONVEYOR
	.0012	1	PLACE ON CONVEYOR
80	.0050		MOVE TO WASHING OPERATION (CONVEYOR)
	.0037	2	WASH AND DRY HEADS
12	.0030		MOVE TO DIE STAMP TABLE
4	.0007	3	PLACE ON TABLE
	.0014	4	DIE STAMP
4	.0008	5	PLACE ON CONVEYOR
14	.0035		MOVE TO PAINT BOOTH
	.0042		PAINT HEADS AND INSPECT
10	.0025		MOVE TO DRYING SECTION
6	.0015	7	REMOVE FROM CONVEYOR
	.33	2	DRY
	.0015	8	LOAD ON TO PALLETS
30	.0006		MOVE TO R.R. CAR

Fig. 3. Typical flow process chart, material type

A flow process chart represents the sequence of all operations, transportations, inspections, delays, and storages occurring during a process or procedure, and includes information considered desirable for analysis, such as time required and distance moved. It is used to follow either material or men through a process and is an essential tool for material handling studies. Figure 3 shows a flow process chart of the material type.

Operation analysis. The methods engineer studies all major factors which affect a given operation, then consciously and open-mindedly applies the questioning attitude and asks either himself or others why, where, what, when, who, and how. If done systematically, worth-while improvements are almost certain to result.

Operation analysis is sometimes defined as common sense systematically applied. The 10 major points of operation analysis listed in the order in which they are considered are purpose of operation, design of part, process analysis, inspection requirements, material, material handling, work place layout and tool equipment, common possibilities for job improvement, working conditions, and method.

To determine the best way to perform an operation, the methods engineer must know why it is be-

design engineer may have overlooked. An analysis of the process, with or without the aid of process charts, may suggest elimination, combination, or change of operation that will reduce costs. A consideration of inspection requirements will often

...taining necessary quality ...duct is ...netimes leads to economies. Methods for handling the materials are particularly important; a study of them may eliminate a few unnecessary steps or produce a completely new philosophy of manufacturing.

Analysis of work place layout and tools is a fruitful source of savings. The methods engineer studies them in connection with the following common possibilities for job improvement: (1) install gravity delivery chutes, (2) use drop delivery, (3) compare methods if more than one operator is working on same job, (4) provide correct chair for operator, (5) improve jigs or fixtures by providing ejectors and quick-acting clamps, (6) use foot-operated mechanisms, (7) arrange for two-handed operation, (8) arrange tools and parts within normal working area, (9) change layout to eliminate back tracking and to permit coupling of machines, and (10) use improvements developed for other jobs.

After an analysis of the conditions under which the work must be performed, the methods engineer considers all of the ideas for improvement that have occurred to him and combines them into one improved and practical method.

Motion study. Manual and eye movements that occur in an operation or work cycle are studied to eliminate wasted movements and establish a better sequence and coordination of movements.

In making a motion study, the methods engineer studies each individual motion in detail and tries to shorten it, combine it with others, or eliminate it altogether. He identifies ineffective motions and seeks

to replace them with effective motions. As a result of this intensive study, he is almost always able to develop an easier, quicker way of doing the work.

Motions may be studied individually by direct observation or they may be observed and timed in groups during the making of a stop watch time study. Time-study data are useful when studying the relationship between the work done by a man or groups of men and the work done by a machine or groups of machines. When these data are properly arranged on a multiple-activity process chart, possibilities for improvement are often revealed.

Quite effective motion studies can be made by observation with the aid of one of the predetermined elemental time systems (*see* WORK MEASUREMENT). As a tool of motion study, they are similar in ease of application to motion study by observation and, in the hands of one trained in their use, approach the results obtained by micromotion study in developing improved methods.

When more intensive motion study is desirable, the methods engineer turns to micromotion study. He takes a motion picture of the operation under study, and analyzes the resulting film frame by frame. When the results of the analysis are recorded on an operator process chart, he has a clear and permanent record of the method. Motions are classified into therbligs, also called Gilbreth basic elements or basic divisions of accomplishment. The therbligs generally recognized at present are (1) transport empty or reach, (2) transport loaded or move, (3) grasp, (4) position, (5) disengage, (6) release, (7) do, (8) examine, (9) change direction, (10) preposition, (11) search, (12) select, (13) plan, (14) balancing delay, (15) hold, (16) avoidable delay, (17) unavoidable delay, and (18) rest to overcome fatigue. The first eight, in general, accomplish necessary work, the next six retard accomplishment, and the last four do not accomplish. The last two groups should be eliminated when possible.

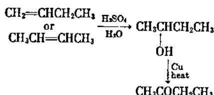
Material handling. Within a factory or other enterprise, handling of material adds to the cost of a product without adding to its salable value. Therefore material handling should be eliminated or minimized. Because material handling is usually an appreciable cost, specialists are sometimes employed to concentrate on such problems. Process charts are useful to give a clear picture of present material handling methods and to point the way to new and improved ones.

Methods design. In designing a new method, the methods engineer is guided by concepts of motion economy. By the use of balanced, two-handed methods and the elimination of all nonproductive and fatiguing therbligs, the engineer designs efficient motion patterns.

To communicate the final approved methods to others, the methods engineer writes a methods specification, describing in as much detail as the importance of the job justifies exactly how it is to be performed. *See* METHODS ENGINEERING. [H.B.M.]

Methyl ethyl ketone

A chemical compound, also called 2-butanone, of formula $CH_3COCH_2CH_3$. It is a liquid with boiling point 79.6°C, melting point −86.9°C, specific gravity 0.805. Methyl ethyl ketone is an important, low-cost (12.5¢ per pound in 1958) solvent and



chemical intermediate. It is manufactured by dehydrogenation or oxidation of 2-butanol (*sec*-butyl alcohol) which comes from the hydration of 1-butene or 2-butene. *See* ACETONE; KETONE. [D.A.S.]

Metric system

A system of units used in scientific work throughout the world and employed in general commercial transactions and engineering applications in most of the civilized world except for Great Britain and the United States. The metric system had its origin in 1791, when a distinguished committee of the French Academy, including J. L. Lagrange and P. S. Laplace, made a report to the National Assembly which was subsequently adopted and put into effect in France. The original report defined a standard of length, the meter, as one ten-millionth (0.0000001) of the earth's meridian quadrant at sea level. In terms of the original system, the mass unit known as the gram became a secondary standard based on the centimeter and the density of water. The original *mètre des archives* was constructed of platinum in 1793. The original geodetic surveys proved inaccurate, and the original length and mass units have been superseded by others which have been accepted by international agreement.

In 1875, the International Bureau of Weights and Measures was established at Sèvres, France. The meter is presently defined as the distance between two fine lines engraved on the international prototype meter bar when the bar is at the temperature of melting ice. Similarly, the mass unit known as the kilogram is defined as the mass of the international prototype kilogram. The prototypes, which are composed of a platinum-iridium alloy, are carefully preserved at Sèvres. *See* KILOGRAM; METER (UNIT).

The chief advantage of the metric system is that it is based on standards that have been accepted by international agreement, and it therefore provides a common basis for all scientific measurements. A second advantage of the metric system lies in the fact that only decimal multiples and submultiples of the fundamental length and mass units and of other derived units are employed. The following prefixes are used to indicate multiples of the meter,

the gram, the second, the ampere, and other units.

deka	for 10	deci	for 10^{-1}
hecto	for 10^2	centi	for 10^{-2}
kilo	for 10^3	milli	for 10^{-3}
mega	for 10^6	micro	for 10^{-6}
giga	for 10^9	nano	for 10^{-9}
tera	*for* 10^{12}	*pico*	*for* 10^{-12}

The time unit employed in the metric system is the second, defined as 1/86,400 of a mean solar day. The mean solar day is the average length of the solar day measured from one passage of the Sun across the meridian to the next, the average being taken over a long period of years. *See* DAY; SECOND (TIME UNIT); UNITS, SYSTEMS OF. [D.WI.]

Mho

A unit of conductance. The conductance of an electrical conductor is the reciprocal of its resistance measured in ohms. The conductance G is defined by the relation

$$I = GV$$

Here I is the current in amperes, V the potential difference in volts, and G the conductance in mhos. *See* OHM; RESISTANCE, ELECTRICAL. [F.J.B.]

Mica

Any one of a group of hydrous potassium-aluminum silicate minerals, some of which also contain Mg, Fe(II), Mn, Li, Fe(III), and Ti as major constituents. All the micas are characterized by perfect, easy basal cleavage (micaceous cleavage). The main series and species are muscovite-lepidolite, phlogopite-biotite [containing Mg and Fe(II)], manganophyllite (Mn mica), paragonite (Na mica), and zinnwaldite (Li-Fe mica). Less common species are roscoelite (V mica), and taeniolite (Li-Mg mica). *See* BIOTITE; LEPIDOLITE; MUSCOVITE; PHLOGOPITE; SILICATE MINERALS; *see also* CHLORITOID.

The micas of economic importance are muscovite and phlogopite, whose technological significance stems from combinations of the following properties: perfect basal cleavage flakes that are flexible, elastic, tough, and translucent to transparent; and low electrical and heat conductivity. Both species have high dielectric strengths, that is, the ability to withstand high voltages without puncture. The power factor of mica, which determines its suitability for use in capacitors, is the ratio of the total power loss in a capacitor in which mica is the dielectric to the total volt-amperes supplied to the capacitor.

Structure. The mica group is structurally complex; muscovite occurs in three distinct structural types, lepidolite in at least four, phlogopite in three, and biotite in at least six, representing stacking variations of successive layers. Most micas are monoclinic pseudohexagonal, but a few are hexagonal or triclinic. Crystals commonly are tabular with prominent basal planes and hexagonal or lozenge-shaped outlines (edges intersecting at nearly 60°, or 120°). The mica structure is based on sheets of linked $(Si,Al)O_4$ tetrahedrons (phyllosilicates), which accounts for the characteristic cleavage into sheets.

Thin cleavage sheets of unaltered micas are both flexible and elastic. Percussion figures may be developed on cleavage plates by striking the surface sharply with a dull-pointed tool to yield a six-rayed star, two lines of which are parallel to the prism edges. The third, most prominent ray is parallel to the trace of the clinopinacoid. Optically, most micas are biaxial and levorotatory, with moderate birefringence. *See* BIREFRINGENCE.

Upon heating in a closed tube, micas yield water; they fuse with difficulty. Specific gravities are in the range 2.8–3.2; hardnesses are between 2 and 3 (Mohs scale).

Occurrence. [illegible] muscovite,

[illegible]

ents in a variety of igneous, metamorphic, and sedimentary rocks, and in many mineral deposits. In igneous rocks, biotite is the most common, appearing in some gabbros, diorites, tonalites, granodiorites, syenites, nepheline syenites, and granites. Phlogopite occurs in some peridotites, and muscovite appears in a few granites. Both muscovite and biotite are widespread and abundant constituents of pegmatites, commonly occurring in crystals 1 in to several feet across, called books. Pegmatites are the chief sources of commercial muscovite mica.

In metamorphic rocks, muscovite and biotite appear chiefly in schists and gneisses, in association with such species as chlorite, garnet, kyanite, staurolite, quartz, and feldspar. Both biotite and muscovite occur as detrital minerals in sands and sandstone, but biotite is chemically altered much more readily than muscovite, and tends to disappear. Muscovite is an exceedingly stable mineral in the weathering and sedimentary environments. Micas formed within the sedimentary environment (authigenic) are apparently chiefly of the hydromica (illite) type in which the oxonium ion (H_3O^+) is substituted isomorphously for the potassium ion, K^+, and thus are species generally regarded as belonging to the clay minerals group.

Muscovite, especially as the fine-grained variety commonly called sericite, is a widespread gangue mineral in many hydrothermal mineral deposits, either in the deposits themselves, or in the adjacent altered wall rocks. The process of sericitization commonly accompanies the development of metalliferous ore deposits formed within the intermediate temperature-pressure range (mesothermal deposits). Sericite also is formed as a common low-temperature replacement mineral (in many instances pseudomorphous) of many species, for example, sillimanite, andalusite, kyanite, cordierite, staurolite, beryl, potash feldspars, plagioclase, tourmaline, and topaz.

Uses. Commercial mica is of two main types: (1) sheet and punch, and (2) scrap or flake. High quality sheet muscovite is essential as the dielectric

in radio and radar circuit capacitors, in magneto-capacitors and coils, and as insulation in radio tubes and airplane spark plugs. Muscovite of lower quality (electric mica) is extensively employed as insulator elements in hot plates, toasters, irons, and other electrical home equipment. Built-up mica plate (micanite), which is made by bonding thin layers of mica splittings together with shellac or glyptal binders and drying under pressure, is extensively used as insulation between copper segments of generators and motors.

Scrap and flake mica is ground, both wet and dry, for use in paints, decorative inks, rubber filler, and coatings on roofing materials and waterproof fabrics. The chief producers of mica are the United States (North Carolina, New England), India, and Brazil; the United States is the largest consumer Synthetic fluorphlogopite is produced commercially as substitutes for natural mica in some applications. *See* CAPACITOR; INSULATION, ELECTRIC. [E.W.H.]

Bibliography: E. W. Heinrich, A. A. Levinson, D. W. Levandowski, and C. H. Hewitt, *Studies in the Natural History of Micas*, Univ. Mich., Eng. Research Inst. Final Rept., Project M978. 1953; E. W. Heinrich, *Microscopic Petrography*, 1956; E. H. Kraus, W. F. Hunt, and L. S. Ramsdell, *Mineralogy*, 5th ed., 1959.

Mica schist

A widely distributed group of rocks of medium to high metamorphic grade, composed essentially of mica and quartz and exhibiting a foliated or schistose structure which is easily revealed by the parallel orientation of the mica flakes. Of lower metamorphic grade and of finer grain are phyllites which contain less biotite but otherwise are similar to mica schists, and pass into it by gradual transitions. *See* METAMORPHIC ROCKS; PHYLLITE.

The simple mineral composition of the ordinary mica schist (quartz and mica only) demonstrates [illegible]

[illegible] all possible ions try to accommodate themselves in its crystalline structure; those that cannot are simply carried away. However, this process cannot go on indefinitely. Eventually the composition is overstrained and if, for example, the alumina concentration is high, new minerals like andalusite, sillimanite, cordierite, and almandine garnet will develop. Rare types contain corundum, paragonite, and fuchsite. Minor accessories are rutile, zircon, tourmaline, and iron ores. When feldspar minerals are abundant, mica schist grades into gneiss; with an increase of quartz it grades into micaceous quartzite.

Mica schists are most widely distributed in Precambrian areas and in the younger eroded mountain ranges where they usually represent metamorphic equivalents of aluminum-rich sediments (clays and argillites). However, tuffs and certain acid igneous rocks (rhyolites, granites) may form part of the primary constituents. Usually metasomatic [illegible]

[illegible] is increased during metamorphism while alumina (Al_2O_3) is lost. [T.F.W.B.]

Micelle

A colloidal particle together with its surrounding stabilizing layer. In the extrinsic sols (stability attributed to electric charge), the term ionic micelle includes the colloidal particle plus its Helmholtz or Goüy double layer. In the intrinsic sols (stability attributed to solvation) at or near the isoelectric point, the term neutral micelle includes the colloidal particle plus its adsorbed layer of stabilizing solvent molecules. See COLLOID; ISOELECTRIC POINT.

Hermann Staudinger used the term in a somewhat different sense to refer to colloidal particles composed of single macromolecules. These are the molecular colloids, as opposed to the micellar colloids which contain colloidal particles consisting of aggregates of numerous small molecules or groups of atoms.

The colloidal particles in colloidal electrolytes, such as the soaps, are commonly referred to as micelles. The concentration at which aggregation occurs to form micelles is referred to as the critical micelle concentration. This concentration may be detected by several experimental methods, including light scattering, electrical conductivity, osmotic pressure, and dye adsorption. *See* SOAP AND DETERGENT.

The extrinsic colloids have been studied from the point of view of identifying the ionic species comprising the inner and outer layers of the double layer. However, confusion has resulted from the attempts of many investigators to consider each micelle to be a complex chemical compound. X-ray diffraction analysis of the bulk structure of the colloidal particle as it flows through an x-ray camera demonstrates that the colloidal particles themselves possess the simple chemical composition one would expect, a situation not obviated by the fact that various stabilizing ions may be adsorbed on their surfaces to produce a stabilizing electrical double layer.

In the colloidal electrolytes, the micelle is considered to consist of an aggregate of numerous simple ionic species. Although at low concentration, the micelle structure may approximate the Hartley spherical model, in more concentrated solutions the laminar model prevails, as evidenced by streaming double refraction and x-ray diffraction measurements. [W.O.M.]

Microbar

A unit of sound pressure. One microbar is equal to 1 dyne/cm², or 0.1 newton/m². A reference level of 1 microbar is quite commonly used in the calibration of microphones, hydrophones, and loudspeakers. See SOUND PRESSURE. [W.J.C.]

Microbiological methods

For controlled study of microorganisms, the microbiologist must have at his disposal adequate methods for quantitatively enumerating, isolating, identifying, and maintaining cultures in pure form. To achieve these objectives, he must have available sterilized culture media, glassware, and other equipment.

The majority of microorganisms can be cultivated in so-called synthetic media whose chemical composition is known. Also available are a large number of commercially prepared dehydrated media which contain partially hydrolyzed animal or vegetable protein, extracts of meats or yeast, and an energy source, usually a fermentable carbohydrate. Rehydration and sterilization of these prepared media in most instances furnish adequate growth media for the purpose of identifying or determining the physiological and biochemical characteristics of the microorganism under study. Some few microorganisms, for example, viruses and rickettsiae, cannot be cultivated in artificial media, and such techniques as animal inoculation, and tissue and embryonated egg culture must be used for their maintenance. Serologic techniques may also be used for identification and quantitation of some microorganisms.

Because of their small size, bacteria can be examined only with a microscope. And because their refractive index is close to that of water, the contrast between organisms and their environment is not very great. In order to increase this contrast, and thus facilitate microscopic examination, the cells are often stained by artificial means. Observations in the living state have, however, many distinct advantages. *See* CULTURE TECHNIQUE; EFFECTIVE DOSE 50; INFECTIVE DOSE 50; LETHAL DOSE 50; SEROLOGY; STAIN (MICROBIOLOGICAL); STERILIZATION. [C.F.N.]

Microbiology

The science and study of microorganisms. These include the protozoa, algae, fungi, bacteria, viruses, and rickettsiae.

Because of their small size, microorganisms cannot be seen without the aid of microscopes. They were discovered in the seventeenth century by Antony van Leeuwenhoek (1632–1723), a Dutch draper and microscopist, who found them in many different environments. He described various types, many so clearly that they can be recognized today.

The development of microbiology has been uneven. During the second half of the nineteenth century, it was recognized that some types of microbes cause diseases of animals and plants. This led to an early emphasis on the medical and sanitary aspects of [illegible] studied large[illegible] Another aspe[illegible] the part playeu by microbes in maintaining the cycle of matter on earth. This function implies that they are usually associated with the spoilage of food products and with transformations in water and soil. Some microbes were found to produce substances of importance to man, and they became important industrially.

Thus, microbiology at first developed largely as an applied science, with little attention paid to the essential biological properties of the microorganisms themselves. However, the studies on their physiology and biochemistry, carried out in the 1920s, pointed to the fundamental unity underlying the behavior of all living organisms. The enormous diversity encountered among the microbes and the specificity of their physiological patterns have helped make these organisms favorite objects for biochemical investigations. These investigations, aided by the available methodology for rigorous experimental control, have contributed greatly to the rapid advances made during the past few decades in the understanding of biochemical phenomena. This, in turn, has stimulated an interest in all sorts of microbes, with the result that microbiology is becoming acknowledged as an important branch of biological science.

The rapid growth of microbes and the enormous numbers that can easily be handled in a small space have also led to their use for studies in heredity and genetics. Many spectacular achievements have been scored in this field.

For medical and sanitary aspects of microbiology, *see* BACTERIOLOGY, MEDICAL; EPIDEMIOLOGY; IMMUNOLOGY; MYCOLOGY, MEDICAL; PARASITOLOGY, MEDICAL; RICKETTSIOSES; VIRUS. For food products spoilage, see FOOD MICROBIOLOGY; MILK. For microbial transformation in water and soil, *see* SOIL MICROBIOLOGY; WATER MICROBIOLOGY. For industrial aspects, *see* INDUSTRIAL MICROBIOLOGY. For microbial biochemical aspects, *see* BACTERIAL PHYSIOLOGY. For methodology, *see* MICROBIOLOGICAL METHODS. For taxonomy, see BACTERIA, TAXONOMY OF; FUNGI. [C.B.V.N.]

Bibliography: W. Braun, *Bacterial Genetics*, 1953; W. Bulloch, *The History of Bacteriology*, 1938; D. G. Catcheside, *The Genetics of Microorganisms*, 1951; C. Dobell, *Antony van Leeuwenhoek and His Little Animals*, 2d ed., 1958; A. J. Kluyver and C. B. van Niel, *The Microbe's Contribution to Biology*, 1956; E. Large, *The Advance of the Fungi*, 1940; R. Y. Stanier, M. Doudoroff, and E. A. Adelberg, *The Microbial World*, 1957; K. V. Thimann, *The Life of Bacteria*, 1955.

Microcircuitry

Electronic circuitry, several orders of magnitude smaller than that used in conventional electronic equipment. Rather than being built up from individual electronic components, the microcircuit is usually fabricated as an integral whole.

One form of microcircuit may be essentially a conventional circuit, highly miniaturized through the use of special fabrication processes, for example, high-vacuum thermal evaporation. This type of



resentatives of the Myxomycetes, the slime molds
temonitis axifera.
Hemitrichia stipitata.
omatricha typhoides.
Hemitrichia abistina.
Physarum viride.







applications of electronic circuitry also are becoming more and more critical and complex, and in many cases a major gain in reliability is needed before commercial installations can be profitable. *See* MINIATURIZATION OF EQUIPMENT. [D.W.M.]

Microcline

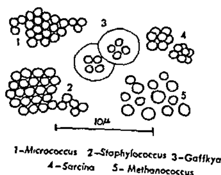
The name for a triclinic potassium-rich feldspar. The microcline structure is built up of rather pure $KAlSi_3O_8$. Most of any $NaAlSi_3O_8$ found in the bulk composition of microcline crystals is present as exsolved albite (frequently up to 25 mole %) in the form of microcline-perthite. Microcline is usually twinned with typical cross hatching indicating that it usually grew (stably or metastably) as sanidine (monoclinic) and subsequently changed during geological time into the triclinic modification by passing through a state called normal orthoclase. It can invert into sanidine under prolonged heat treatment at high temperature; however, all attempts to effect the change from sanidine to microcline in the laboratory or to synthesize microcline have been unsuccessful. This is explained by the fact that microcline has an ordered Al/Si distribution in contrast to the disordered one in sanidine, and that the ordering process is a very sluggish one at low temperature. Microcline is typically found in granites and pegmatites. Green microcline is known as amazon stone or amazonite. *See* FELDSPAR; GEM; IGNEOUS ROCKS; PERTHITE. [F.LA.]

Micrococcaceae

A family of bacteria of the order Eubacteriales

organotrophic; that is, growth is dependent upon oxidation or fermentation of exogenous organic substances. Motility and endospore formation are rare; some species are brightly pigmented. The genera of this family are as follows.

1. *Micrococcus.* The cells are arranged in irregular masses; metabolism is strictly oxidative. Some species form red, yellow, or orange pigments. The



Some genera of the Micrococcaceae. (V. B. D. Skerman)

micrococci are found in a great variety of habitats. Most species are saprophytic, but a few may be parasitic, although not pathogenic.

2. *Staphylococcus.* The cells are arranged in irregular masses; the metabolism is facultatively anaerobic. One species, *S. aureus*, is one of the most important pathogens of man, causing, among other things, boils, abscesses, surgical infections, and food poisoning. *See* FOOD POISONING, BACTERIAL; STAPHYLOCOCCUS.

3. *Gaffkya.* The cells occur in the animal body and in special media as tetrads. In ordinary culture media, they occur in pairs and irregular masses. The metabolism is aerobic to facultatively anaerobic. The species are parasitic and pathogenic for mice, guinea pigs, and lobsters.

4. *Sarcina.* The cells are arranged in cubical packets; the metabolism may be strictly aerobic to obligately anaerobic. Aerobic, pigmented species are common in air. Anaerobic species, found in soil and other habitats, may carry out an alcoholic, butyric, or methane fermentation. *See* FERMENTATION.

5. *Methanococcus.* The cells are arranged in irregular masses; the organisms are strictly anaerobic, fermenting various organic compounds with the production of methane. Species are found in habitats such as soil, sewage, and mud. Along with other methane bacteria, they play an important role in nature in the anaerobic mineralization of organic matter. *See* SOIL MICROBIOLOGY.

6. *Peptococcus.* The cells are arranged in irregular masses; the organisms are strictly anaerobic, fermenting a variety of organic compounds with the production of various products, but never methane. Most species have been isolated from normal and diseased areas of the human body, where they exist as parasites rather than pathogens. *See* EUBACTERIALES. [H.C.D.]

Microfilming

The technique of making a photographic film record on reduced scale of documents, printed matter, and the like, which can be enlarged for reading. Reductions may range from 8 to 40 times, although in extreme cases a reduction of 60 times has been used. Microfilming is a type of microphotography. For details, *see* PHOTOCOPYING PROCESSES; PHOTOGRAPHY. [W.C.]

Bibliography: G. W. W. Stevens, *Microphotography*, 1957.

Micromanipulation

The techniques and science of microdissection, microvivisection, microisolation, and microinjection. This field of knowledge is also known as micrurgy. Almost any mechanical operation can be performed under high magnification by micrurgy when properly instrumented. Applications of micrurgy are many, especially in biology, cancer research, microchemistry, biophysics, colloid technology, and metallurgy.

Applications. The applications of micrurgy to the study of living cells include single-cell isolation, especially of bacteria, spores, and ascites tumor cells; microdissection of virus inclusion bodies in plant cells and the neuromotor system in ciliated protozoa; measuring cell adhesiveness; enucleating cells; stretching chromosomes; microinjecting pH indicators, salt solutions, enzymes, drugs, and oils into cells; bioelectrical studies involving action and membrane potentials; electrical resistance of protoplasm; properties of extraneous coats; mechanisms of cell division; cytochemistry with the isolation of subcellular structures; and subcellular transplantations. Subcellular structures that have been successfully transplanted from cell to cell include micronuclei of ciliated protozoa, normal and irradiated cytoplasm, nuclei of amebas, nuclei of differentiated frog cells into cytoplasm of unfertilized eggs, inclusion bodies of viral origin, and normal or malignant nucleoli usually with attached nucleolar chromosomes. *See* AXENIC CULTURE; BACTERIA; BACTERIAL ENDOSPORES; BIOPOTENTIALS AND ELECTROPHYSIOLOGY; [illegible]

[illegible] demonstra- [illegible] and of synthetic or natural fibers, measuring consistency of colloids, manufacture and assembly of fine mechanical devices and electronic components such as transistors, orienting fine particles and preparing specimens for electron microscopy, and isolating and determining chemical properties of rare elements such as plutonium. *See* COLLOID; ELASTICITY; MICROSCOPE, ELECTRON; PLUTONIUM; TRANSISTOR.

Instrumentation. The basic instruments required for micrurgy are (1) micromanipulators or micropositioners which convert the relatively crude hand movements to exquisitely fine movements of the microtools; (2) microtools such as needles, pipes, hooks, loops, electrodes, scalpels, or forceps; and (3) microscopes and accessories.

Over 200 different types of micromanipulators have been described since Schmidt (1859) published a description of his "Microscopic Dissector." The superb micropositioners are instruments of great precision with which the microneedle, micropipet, or other microtool may be exactly positioned in the field of a microscope at any magnification. These instruments range from simple rack and pinion assemblies to massive, accurately fitted ball bearing slides actuated by precise feed screws or other drives. The earlier micropositioners were screw operated; however, in recent years there has been a trend toward lever- or joystick-controlled micropositioners. In some, movements may be transmitted from the control lever to the slides by pneumatic or hydraulic means; in others, the movements are directly transmitted through mechanical couplings. One lever-controlled micromanipulator generates motion of microtools through the expansion of electrically heated wires. A micropositioner driven by servomechanisms has been designed and is under construction.

The microtools, such as needles, hooks, loops, and pipets, can be fabricated manually by using a simple gas microburner and glass rods or tubing. Glass is the best substance for making microtools, since this is the only material that consistently has ample rigidity even when reduced to micro, or even submicro, dimensions. Several mechanical devices, usually with electric heating elements, are available for making micropipets. Micropipets with special tips or shapes can be constructed in the microforge. Microforges are optical-mechanical devices for controlling the position of needles or pipets in the field of a low-power microscope by a simple micromanipulator. With these devices, microhooks and microloops can be made; micropipets can be bent near the tip, and their minute openings may be beautifully fire-polished. The tip of a microneedle may be submicroscopic in size ($<0.2\ \mu$). Micropipets are frequently used with functional tips less than 0.5 μ in diameter. Microelectrodes with tips of the order of 1 μ are routinely made and used in biophysical laboratories *See* BIOPHYSICS.

The glass microtools are mounted in metal holders which, in turn, are held by clamps on the micropositioners. The better instruments provide coarse adjustments for the preliminary centering and orientation of the microtools in the field of the microscope. Micropipets usually require special holders which are connected to the microinjectors by flexible metal or plastic tubing.

Any conventional research microscope equipped with a good mechanical stage and a long-working-distance substage condenser will suffice for most micrurgical studies. Ordinarily, biological specimens, such as living cells suspended in a fluid medium, are mounted on cover glasses and supported by a box-shaped device, the moist chamber. The moist chamber, which prevents excessive evaporation of the preparations, is positioned by the mechanical stage. With conventional microscopes, the cells are suspended in a hanging drop from the bottom surface of a cover glass. With inverted microscopes, however, lying drops are used so that the cells can rest on the upper surface of the cover glass. *See* MICROSCOPE, OPTICAL.

Procedures. The basic micrurgical procedures include setting up micromanipulator, microscope, and light source; preparation of moist chamber; centering of microtools in optical field; proper vertical placement of microtools in relation to depth of hanging drop or height of lying drop; preparation and placement of cells or other material on cover glasses and subsequent mounting on the moist chamber; practice in manipulating needles or pipets in the field of the microscope by familiarizing oneself with all the adjustments provided by the micromanipulators; practice in manipulating microinjectors; replacement of damaged or dirty microneedles or micropipets; chan

… from low to high powers and vice

a cell, such as nucleoli, chromosomes, and mitochondria, can be transplanted into another cell with a micropipet by microinjection techniques. Such microsurgery requires precise micrurgical and microinjection instrumentation. The tips of the micropipets must be exactly positioned towards the part of a cell to be removed, and this can be easily accomplished with excellent micropositioners. Microinjections can be executed properly only if volume displacement in the microinjector is under control. Heretofore, such controls were provided by fine pistons (0.005-in. diameter) actuated by complex, carefully fitted, worm-gear-feed screw mechanisms. A new method of controlling minute volume changes was recently developed by employing the principle of the differential piston. The pistons, about 1 mm in diameter, with one piston slightly larger than the other, are mounted on a common carrier bracket. Each piston enters the same volume chamber, but from opposite ends. As one piston is moved into the chamber, the other piston moves out. Accordingly, the volume displaced per unit length of travel will be the difference in volume displaced by the two pistons. The piston actuator is a slightly modified coarse-fine focusing mechanism such as is used on microscopes. This unit can be driven by a servomotor, thereby establishing a programmed or semiautomatic microinjector.

Leitz micromanipulator. One of several recent developments in micrurgical instrumentation and procedures and the related equipment is shown in Fig. 1. Included are four Leitz micromanipulators mounted on a special stand along with four microinjectors. Also mounted on the same stand is a video camera for closed-circuit television.

The Leitz micromanipulator consists of three ball-bearing slides mounted at right angles to one another to produce movements in three directions in space, for example, north-south, east-west, and up-down. A single joystick lever controls all horizontal fine motions. Vertical movement is provided

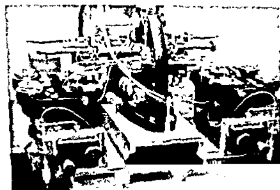


Fig. 1. The Leitz micromanipulator.

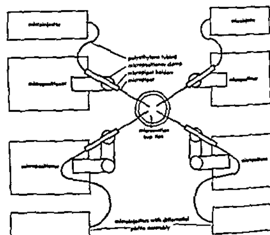


Fig. 2. A top view of the Leitz micromanipulator shown as a block diagram.

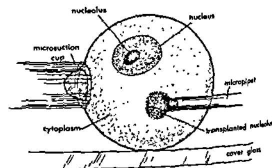


Fig. 3. Diagram of a cell held by a microsuction cup with nucleolus being transplanted.

by a coarse-fine adjustment operated by turning two coaxial knobs. There are also two horizontal coarse controls for preliminary positioning of the microtools. Movement is transmitted from the joystick to the two horizontal slides through a ball-sphere segment. By raising or lowering the eccentrically mounted ball segment, the ratio between lever movement and needle motion can be continuously varied from 16:1 to 800:1.

The four micropositioners are grouped around the microscope (Fig. 2).

Each micropositioner carries a micropipet, a micropipet holder, and micropositioner clamp. Two microinjectors mounted behind the micropositioners are connected to the micropipet holders by polyethylene tubing. Two of the microinjectors consist of glass syringes, the spring-loaded pistons being moved by micrometer screws. A modified micrometer with a small piston provides the necessary fine volumetric control. The micropipets connected to these microinjectors have microsuction cup tips.

In front of the other two micropositioners, two microinjectors of the differential piston type are mounted. These are also connected to micropipet holders by polyethylene tubing and are used for transplanting nucleoli and other subcellular structures. These microinjectors also include a large

piston moved by a micrometer screw for coarse volume adjustments, in addition to the differential piston assembly, which is actuated by a coarse-fine focusing mechanism.

Figure 3 illustrates a cell held by a microsuction cup. A micropipet is also shown inside the cell, with a nucleolus partly extruded into the cytoplasm. The micropipet tip is precisely positioned inside the cell by a micropositioner. The nucleolus is implanted into the cytoplasm by a microinjector. The spherical cell is held by the microsuction cup, which is connected to and operated by the microinjector. A side view with the cell resting on the upper surface of a cover glass is also shown. The objective of an inverted microscope would be placed below the cell. Since the cover glass is the only obstacle between the cell and the objective, it is possible to use high-powered objectives even though their working distances may be short. A conventional microscope equipped with long-working-distance objectives could also be used in this instance.

The transplantation of nucleoprotein-rich subcellular structures, along with the induction of various cell changes, can add new horizons to the study of somatic cell genetics. The chance inclusion into cells of, or infection by, some exogenous subcellular particle is no longer a necessity. A selected structure from one cell may be deliberately removed and placed precisely in another cell. The interaction of viruses and cells, especially where striking cellular effects may be induced, such as interference with normal processes of differentiation, could very well be one of the most fruitful areas of future work. Much of the success in survival and propagation of the cells following subcellular transplantation will be enhanced by more precise procedures as well as by the most modern [illegible]

[M.J.K.]

[illegible] (eds.), *Tl*[illegible] [illegible]1.), *McClung's Handbook of Microscopical Technique*, 3d ed., 1950; M. J. Kopac, Cytochemical micrurgy, *Intern. Rev. Cytol.*, 4:1–29, 1955.

Micrometeorology

A branch of atmospheric dynamics and thermodynamics. Micrometeorology typically deals with the transport (or flux) of air properties in the vertical direction (conduction and convection) and the vertical variation of these fluxes. Micrometeorology deals primarily with the interaction between atmosphere and ground; the interchange of masses, momentum, and energy at the earth-air interface; in short, with the lower boundary conditions of atmospheric processes. Thus, there are strong interconnections with those sciences that deal with the media underlying the atmosphere (*see* SEA WATER; SOIL; *see also* OCEAN-METEOROLOGICAL RELATIONS). Micrometeorology, along with other branches of meteorology, is also concerned with investigating and predicting the transport and dispersion of pollution from such sources as smokestacks and automotive exhausts in the lower atmosphere.

Scale of focus. Micrometeorology differs from general meteorology (in particular, meso- and macrometeorology) in scale and with respect to the features of the atmosphere studied (*see* METEOROLOGY). All of the various branches of meteorology deal with the temporal-spatial variations of the weather elements (such as air density, temperature, components of momentum, and mixing ratios). Meso- and macrometeorological studies are typically based on standard synoptic observations from an internationally organized network (with 100 km as the order of station distance). Of major interest are air movements and the transport of little-changing air properties (such as potential temperature, absolute vorticity, mixing ratios) in horizontal directions over large distances, together with the effects of horizontal divergence of air.

Many processes occur in the atmosphere which are incompletely covered by the synoptic network. Examples of such small-scale processes are mountain and valley circulations, sea breezes, and air flow past mountains and islands. The prefix micro is justified by the fact that the detailed study of the temporal-spatial variations of air properties in the lower atmosphere requires close spacing of instruments of relatively small size and low inertia (lag time).

Instrumental needs. The development and use of nonstandard mast-supported thermometers, anemometers and other sensing instruments for the measurement of mean vertical profiles and gradients of temperature, wind speed, and other air properties is a basic requirement of micrometeorological research. Further miniaturization of sensing instruments is necessary for the study of fluctuations of the meteorological elements.

Another basic requirement is the development and use of instruments for the direct measurement of boundary fluxes (such as evaporation, surface stress, and energy transfer, including net radiation). Micrometeorological studies are frequently concerned with the diurnal cycle of heating and cooling, the hydrologic cycle, and the energy dissipation in large-scale air currents due to surface stress (friction).

Under clear skies approximately 80% of solar radiation reaches the ground (20% being absorbed in the atmosphere) where it is partly reflected ("albedo," usually less than 15%) and partly [illegible]

and light, creating irregular scattering and refraction phenomena which account for optical mirages,

magnification from low to high powers and vice versa; and skill in making microtools.

Differential pistons. Many parts of a cell, such as nucleoli, chromosomes, and mitochondria, can be transplanted into another cell with a micropipet by microinjection techniques. Such microsurgery requires precise micrurgical and microinjection instrumentation. The tips of the micropipets must be exactly positioned towards the part of a cell to be removed, and this can be easily accomplished with excellent micropositioners. Microinjections can be executed properly only if volume displacement in the microinjector is under control. Heretofore, such controls were provided by fine pistons (0.005-in. diameter) actuated by complex, carefully fitted, worm-gear-feed screw mechanisms. A new method of controlling minute volume changes was recently developed by employing the principle of the differential piston. The pistons, about 1 mm in diameter, with one piston slightly larger than the other, are mounted on a common carrier bracket. Each piston enters the same volume chamber, but from opposite ends. As one piston is moved into the chamber, the other piston moves out. Accordingly, the volume displaced per unit length of travel will be the difference in volume displaced by the two pistons. The piston actuator is a slightly modified coarse-fine focusing mechanism such as is used on microscopes. This unit can be driven by a servomotor, thereby establishing a programmed or semiautomatic microinjector.

Leitz micromanipulator. One of several recent developments in micrurgical instrumentation and procedures and the related equipment is shown in Fig. 1. Included are four Leitz micromanipulators mounted on a special stand along with four microinjectors. Also mounted on the same stand is a video camera for closed-circuit television.

The Leitz micromanipulator consists of three ball-bearing slides mounted at right angles to one another to produce movements in three directions in space, for example, north-south, east-west, and up-down. A single joystick lever controls all horizontal fine motions. Vertical movement is provided



Fig. 1. The Leitz micromanipulator.

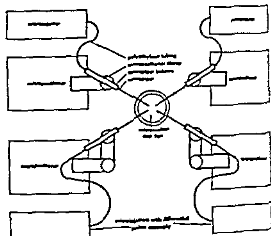


Fig. 2. A top view of the Leitz micromanipulator shown as a block diagram.

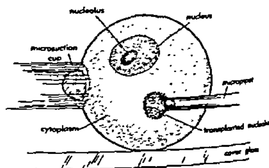


Fig. 3. Diagram of a cell held by a microsuction cup with nucleolus being transplanted.

by a coarse-fine adjustment operated by turning two coaxial knobs. There are also two horizontal coarse controls for preliminary positioning of the microtools. Movement is transmitted from the joystick to the two horizontal slides through a ball sphere segment. By raising or lowering the eccentrically mounted ball segment, the ratio between lever movement and needle motion can be continuously varied from 16:1 to 800:1.

The four micropositioners are grouped around the microscope (Fig. 2).

Each micropositioner carries a micropipet, a micropipet holder, and micropositioner clamp. Two microinjectors mounted behind the micropositioners are connected to the micropipet holders by polyethylene tubing. Two of the microinjectors consist of glass syringes, the spring-loaded pistons being moved by micrometer screws. A modified micrometer with a small piston provides the necessary fine volumetric control. The micropipets connected to these microinjectors have microsuction cup tips.

In front of the other two micropositioners, two microinjectors of the differential piston type are mounted. These are also connected to micropipet holders by polyethylene tubing and are used for transplanting nucleoli and other subcellular structures. These microinjectors also include a large

piston moved by a micrometer screw for coarse volume adjustments, in addition to the differential piston assembly, which is actuated by a coarse-fine focusing mechanism.

Figure 3 illustrates a cell held by a microsuction cup. A micropipet is also shown inside the cell, with a nucleolus partly extruded into the cytoplasm. The micropipet tip is precisely positioned inside the cell by a micropositioner. The nucleolus is implanted into the cytoplasm by a microinjector. The spherical cell is held by the microsuction cup, which is connected to and operated by the microinjector. A side view with the cell resting on the upper surface of a cover glass is also shown. The objective of an inverted microscope would be placed below the cell. Since the cover glass is the only obstacle between the cell and the objective, it is possible to use high-powered objectives even though their working distances may be short. A conventional microscope equipped with long-working-distance objectives could also be used in this instance.

The transplantation of nucleoprotein-rich subcellular structures, along with the induction of various cell changes, can add new horizons to the study of somatic cell genetics. The chance inclusion into cells of, or infection by, some exogenous subcellular particle is no longer a necessity. A selected structure from one cell may be deliberately removed and placed precisely in another cell. The interaction of viruses and cells, especially where striking cellular effects may be induced, such as interference with normal processes of differentiation, could very well be one of the most fruitful areas of future work. Much of the success in survival and propagation of the cells following subcellular transplantation will be enhanced by more precise procedures as well as by the most modern instrumentation. [M.J.K.]

Bibliography: J. Brachet and A. Mirsky (eds.), *The Cell*, vol. 1, 1959; R. M. Jones (ed.), *McClung's Handbook of Microscopical Technique*, 3d ed., 1950; M. J. Kopac, Cytochemical micrurgy, *Intern. Rev. Cytol.*, 4:1–29, 1955.

Micrometeorology

A branch of atmospheric dynamics and thermodynamics. Micrometeorology typically deals with the transport (or flux) of air properties in the vertical direction (conduction and convection) and the vertical variation of these fluxes. Micrometeorology deals primarily with the interaction between atmosphere and ground; the interchange of masses, momentum, and energy at the earth-air interface; in short, with the lower boundary conditions of atmospheric processes. Thus, there are strong interconnections with those sciences that deal with the media underlying the atmosphere (see SEA WATER; SOIL; see also OCEAN-METEOROLOGICAL RELATIONS). Micrometeorology, along with other branches of meteorology, is also concerned with investigating and predicting the transport and dispersion of pollution from such sources as smokestacks and automotive exhausts in the lower atmosphere.

Scale of focus. Micrometeorology differs from general meteorology (in particular, meso- and macrometeorology) in scale and with respect to the features of the atmosphere studied (see METEOROLOGY). All of the various branches of meteorology deal with the temporal-spatial variations of the weather elements (such as air density, temperature, components of momentum, and mixing ratios) Meso- and macrometeorological studies are typically based on standard synoptic observations from an internationally organized network (with 100 km as the order of station distance). Of major interest are air movements and the transport of little-changing air properties (such as potential temperature, absolute vorticity, mixing ratios) in horizontal directions over large distances, together with the effects of horizontal divergence of air.

Many processes occur in the atmosphere which are incompletely covered by the synoptic network. Examples of such small-scale processes are mountain and valley circulations, sea breezes, and air flow past mountains and islands. The prefix micro is justified by the fact that the detailed study of the temporal-spatial variations of air properties in the lower atmosphere requires close spacing of instruments of relatively small size and low inertia (lag time).

Instrumental needs. The development and use of nonstandard mast-supported thermometers, anemometers and other sensing instruments for the measurement of mean vertical profiles and gradients of temperature, wind speed, and other air properties is a basic requirement of micrometeorological research. Further miniaturization of sensing instruments is necessary for the study of fluctuations of the meteorological elements.

Another basic requirement is the development and use of instruments for the direct measurement of boundary fluxes (such as evaporation, surface stress, and energy transfer, including net radiation). Micrometeorological studies are frequently concerned with the diurnal cycle of heating and cooling, the hydrologic cycle, and the energy dissipation in large-scale air currents due to surface stress (friction).

Under clear skies approximately 80% of solar radiation reaches the ground (20% being absorbed in the atmosphere) where it is partly reflected ("albedo," usually less than 15%) and partly [illegible] into sensible heat. Energy striking the [illegible]

a surprisingly complex and variable structure. Strong fluctuations and gradients of temperature, density and moisture affect the propagation of sound and light, creating irregular scattering and refraction phenomena which account for optical mi[illegible]

shimmer, and "boiling," as well as "ducting" of radio and radar beams.

Significant applications. Micrometeorological research is important in that it supplies detailed information about the physical processes in the region of the atmosphere where life is most abundant. It closes the gaps of information from synoptic networks. It produces results useful for applications in various fields such as climatology, oceanography, soil physics, agriculture, biology, chemical warfare, and air pollution (*see* CLIMATOLOGY; INDUSTRIAL METEOROLOGY). Among the branches of atmospheric physics, micrometeorology is the one whose subjects are most amenable to fairly complete experimental description, and to testing of the theoretical models. The characteristic scale is small enough that it is both feasible and economical to attempt control of natural processes in the lower atmosphere, for example, by artificial changes of albedo or other surface characteristics, such as roughness and windbreaks; by irrigation; by the use of smoke screens; by heat supply; and by artificial stirring of air. *See* WEATHER MODIFICATION.

Turbulence research and applications. The mechanism of the vertical flux is essentially one of eddy mixing or turbulent exchange of air properties. It is convenient to distinguish two methods of approach: one deals with the mean vertical gradients caused by turbulence, while the other is based on statistical treatment of turbulent fluctuations recorded by low-inertia (fast-response) instruments. The connecting link between the two approaches is presented by the Reynolds expression of the mean flux of an air property as the average covariance of the fluctuations of vertical wind speed and the property considered.

A powerful tool for research is the spectrum analysis of fluctuations by which the distribution of the variance of a meteorological element over the various frequencies is measured. According to J. Van der Hoven the power spectrum (in the range 0.001–1000 cycles per hour) of horizontal wind speed recorded at the upper levels of the meteorological tower at Brookhaven, Long Island, shows two major peaks of energy, one at about 1 cycle per 4 days, another at about 1 cycle per minute. A broad minimum of energy at frequencies from about 1 to 10 cycles per hour seems to exist under varying terrain and synoptic conditions. This spectral gap appears to separate objectively the macrometeorological from the micrometeorological scale in the atmosphere.

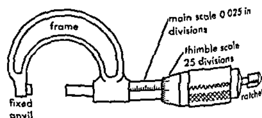
Owing to the relatively high Reynolds number of atmospheric flow, all air motions are turbulent, which means that individual particles do not describe straight and parallel trajectories (as in truly laminar flow) even though the originating force (pressure gradient force) may be constant and uniform and the ground smooth and flat. Turbulence is sometimes visible in the shapes assumed by smoke, and felt in the variable pressure and cooling power of the wind. The intensity of atmospheric turbulence (as measured by the ratio: standard deviation of a wind component divided by mean wind speed) is normally between 0.2 and 0.4, which is significantly larger than that of wind tunnel turbulence. Low-level turbulence derives its energy basically from large-scale air motions and depends on the surface roughness. This mechanical turbulence is intensified by surface heating and damped by nocturnal cooling. Individual gusts can be ascribed to a disturbance, or eddy, of a certain geometrical dimension. There is a wide range of eddy sizes (eddy spectrum). In the vicinity of the ground, eddy sizes increase in proportion to distance from the ground, then decrease with further increase in height.

Atmospheric turbulence affects the flight of airplanes and ballistic missiles (rough air). It is especially troublesome when intensified by surface heating (thermal turbulence) and additional buoyancy forces generated by latent heat release in updrafts that ascend above the condensation level (cumulus convection). There are many features of turbulence research that are of common interest to micrometeorology and aeronautics. [H.H.L.]

Bibliography: R. Geiger, *The Climate Near the Ground*, rev. ed., 1957; H. Lettau and B. Davidson (eds.), *Exploring the Atmosphere's First Mile*, 2 vols., 1957; O. G. Sutton, *Micrometeorology*, 1953; J. Van der Hoven, Power spectrum of horizontal windspeed in the frequency range from 0.0007 to 900 cycles per hour, *J. Meteorol.*, 14:160, 1957.

Micrometer

An instrument for measuring minute distances. For example, in an astronomical telescope, fine hairs or threads from a spider's web are stretched on a movable frame across the focal plane; the frame moves on a rack and pinion or other slow-motion drive with a calibration. The thread is moved from one side of an image to the other and the distance as shown on the calibrated scale is read to measure the apparent diameter of the object, thus serving as a micrometer. A similar device may be used in a microscope.



Machinist's caliper with micrometer adjustment. (R. J. Sweeney, Measurement Techniques in Mechanical Engineering, Wiley, 1953)

In machine shops, a caliper is equipped with a fine-screw thread, movable spindle, and fixed anvil for the precise measurement of distances. A micrometer caliper as illustrated is often used to measure differences of 0.0001 in. [F.H.R.]

Micron

A unit of length equal to one-millionth (0.000001) of a standard meter or a thousandth of a millimeter; it is symbolized by the Greek letter μ and is commonly used as a unit for measuring short distances. For example, the wavelengths of infrared radiation range from 1 μ to several hundreds. Further subdivision of the micron by 1000 defines the millimicron (mμ), which is used in spectrophotometry and generally in expressing wavelengths of visible and ultraviolet radiation that are measured to not more than three significant figures. *See* METER (UNIT). [W.F.M.]

Micropaleontology

The branch of paleontology concerned with fossils which are usually studied and identified by means of the microscope. These organic remains, referred to as microfossils, are generally quite small, ranging in size from a few microns to a few millimeters. Some types of larger fossils, such as lime-secreting algae, stromatoporoids, and bryozoans, are properly included, since they are usually studied and identified by microscopy.

Most major groups of organisms are represented by some type of micropaleontological object preserved in sedimentary rocks. These remains may consist of (1) skeletons or shells of entire microorganisms, such as diatoms or foraminifers; (2) immature or embryonic shells of larger organisms such as corals, mollusks, or brachiopods; (3) individual skeletal elements of larger fossils, including sponge spicules, plant spores and pollens, fish teeth and scales, or the various plates and spines of echinoderms. Microfossils may be composed of calcium carbonate, silica, calcium phosphate, strontium sulfate, chitin, or various complex organic compounds.

Unicellular forms of the kingdom Protista (Fig. 1) include coccoliths, lime-secreting algae, diatoms, dinoflagellates, silicoflagellates, chrysomonadines, tintinnids, foraminifers, as well as chitinozoans and hystrichospherids, whose systematic position in the organic world is uncertain.

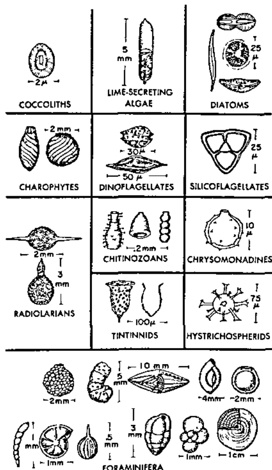


Fig. 1 Microfossils of the kingdom Protista.

Plant microfossils (Fig 2) include pollens, spores, seeds, fossil woods, and leaf cuticles. For detailed information on pollens and spores, *see* PALYNOLOGY; *see also* FOSSIL SEEDS AND FRUITS

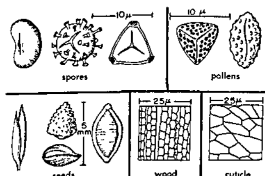


Fig 2. Representative plant microfossils.

Micropaleontological objects representing the animal kingdom (Fig. 3) include skeletal elements of sponges, corals, echinoderms, worms, fishes, mammals; the immature shells of brachiopods, pelecypods, gastropods, and scaphopods, the entire carapaces or shells of ostracods; and toothlike skeletal elements known as conodonts, which have been variously assigned to the worms, fishes, gastropods, and cephalopods.

Occurrence and abundance. Microfossils occur in most types of marine and nonmarine sedimentary rocks, particularly in limy shales, limestones, black shales, siltstones, and fine-grained sandstones. They are usually disseminated through the rock in quantities ranging from a few to several hundred per gram of rock. Certain rock types, however, are composed almost entirely of microfossils. Diatomite is a siliceous shale made up of the bivalved shells of diatoms; chalk is a soft variety of limestone composed of the calcareous tests of foraminifers, and of coccoliths; certain limestones of the Pennsylvanian and Permian systems are essentially composed of fusulinid Foramini' the Gizeh limestone of the Eocene of F

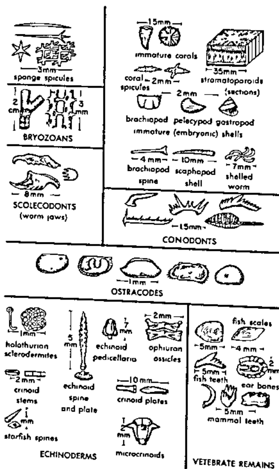


Fig 3. Microfossils of the animal kingdom.

which the Pyramids were built, is made up almost entirely of the shells of the large foraminifer *Nummulites*. Many reef limestones in the sedimentary column are built by lime-secreting algae, corals, bryozoans, and stromatoporids. The dense siliceous rock known as chert is, in many instances, believed to have been formed from accumulations of the shells of diatoms and radiolarians, while plant spores and pollens are believed to be the essential particles composing the rock known as cannel coal.

Collection and preservation. Surface exposures of sedimentary rock are sampled for micropaleontological analysis by collecting about a pint of fresh unweathered rock at the outcrop. Choice of vertical sampling interval depends upon the total thickness of rock exposed and the degree of change in the sedimentary characteristics within the vertical section.

Selection of samples containing microfossils larger than ½ millimeter (mm) may be greatly facilitated by inspecting the sample with a 10 or 15× hand lens. The presence of smaller microfossils cannot be detected in the field without ...

...ngs and from cores obtained in the course of drilling wells for oil, gas, or water. They consist of small chips or rock fragments washed from the bore hole during drilling operations, or of cylindrical cores of the rock strata obtained by the use of a special drilling bit designed for the purpose.

Laboratory treatment to free the microfossils from the enclosing rock matrix varies greatly, depending upon the hardness, texture, and mineral composition of the rock, and upon the size and chemical composition of the microfossils. Since most microfossils are quite fragile, the sample is given the minimum of mechanical and chemical treatment to reduce the rock to individual particles and to liberate the microfossils.

Soft shales, siltstones, and chalks are usually soaked and gently boiled in water containing a deflocculant. This treatment reduces the sample to a thick mud which is flushed through a series of graduated screen sieves to remove the finest particles. The various-sized fractions are dried and placed in envelopes or vials to await inspection with the microscope. Tougher and more resistant shales are broken down by crushing, vigorous boiling with detergents, heating and quenching, freezing and thawing, or by placing an aqueous suspension of the sample in a supersonic field.

Limestones and other carbonate rocks are crushed to pea-sized fragments and the sample is split. One fraction is treated as described above; the other fraction is digested with dilute hydrochloric acid, or with glacial acetic acid, to dissolve the carbonate minerals, leaving a residue of insoluble mineral grains and such noncalcareous microfossils as siliceous sponge spicules, conodonts, diatoms, radiolarians, and plant microfossils.

Disaggregation of coal and shales to liberate spores, pollens and other very small microfossils requires special treatment. This includes digestion of the rock sample with Schulze's solution (a mixture of nitric acid and potassium chlorate) or with hydrofluoric acid; concentration of the fine-grained residues by centrifuging, suspension of smears of the residues in glycerin jelly, selective staining, and mounting of the stained material on glass microslides for microscopic examination.

Microfossils larger than ½ mm are separated from the dried and size-sorted residues by scattering the grains on a small black tray, and inspecting them on the stage of a wide-field stereoscopic binocular microscope at magnifications of 10 or 15 diameters. The individual fossils are selected from the residue by picking them up on the moistened and pointed tip of a small red sable water-color brush (size 000) and transferred to a special micropaleontological slide consisting of a 1- by 3-in strip of black cardboard, ruled into 60 or 100 small compartments and coated with a water-soluble adhesive such as gum tragacanth.

Smaller microfossils are studied in smears or suspensions of fine residues mounted on glass microslides and protected with cover slips. Spores, pollens, diatoms, hystrichospherids, dinoflagellates,

silicoflagellates, and microforaminiferans are studied under the light microscope, at magnifications ranging from 100 to 1500. They are also studied by means of the electron microscope, at high magnifications ranging from 600 to 25,000.

Certain large fossils, such as bryozoans, lime-secreting algae, colonial corals, stromatoporids, and large Foraminifera, are prepared for examination by grinding down an oriented fragment of the large fossil to a thin rock slice through which light can be transmitted after it is mounted on a glass microslide with a cover slip. These sections are studied by transmitted light at magnifications ranging from 30 to 500×.

Applied micropaleontology. Most of the more than 1500 micropaleontologists in the United States are employed in the exploration departments of oil companies, where their work significantly contributes to the knowledge of subsurface sedimentary rocks so essential to successful petroleum exploration. Modern petroleum exploration consists of searching the subsurface sequences of sedimentary rocks in various parts of the world for (1) general areas of marine sediments located near the shorelines of the ancient seas in which they were deposited, and (2) localized spots within these broad areas, consisting of subsurface structures, or geological traps. These include anticlinal domes, folds, faults, and unconformities which serve to entrap the oil and gas into small areas within such permeable reservoir rocks as sandstones and limestones.

In searching for ancient shorelines, micropaleontologists examine the well cuttings and cores for microfossils which indicate a certain ancient environment, characterized by certain ranges of water temperature, or depth of water, or distance from shore. Many species of Foraminifera are very helpful in this regard, particularly those bottom-dwelling forms whose similarity to modern species warrants the assumption that the fossil forms dwelt in about the same environments in the sea as their modern counterparts. Another procedure of diagnostic value, in reconstructing ancient sedimentary environments, is to make quantitative measurements of fossil spores and pollens from the same stratum in a number of wells, and by comparing the numbers of spores and pollens of land-dwelling plants from well to well, to determine the general direction and extent of the ancient land mass.

Determination of the geologic age of a given formation or sedimentary rock unit, and correlation of sequences of sediments from well to well, are important phases of mapping subsurface structures which may localize oil accumulations. Age determination is accomplished by means of guide fossils, which are species of wide geographic distribution and of short vertical, or time, range (see INDEX FOSSIL). Correlation of strata is accomplished by marker horizons. These consist of either a distinctive and easily recognizable stratum of wide horizontal extent, or a zone of guide fossils. In such oil-producing areas as the Texas-Louisiana Gulf Coast, and in southern California, thick sequences of Tertiary marine strata exhibit little or no contrast in rock types or characteristics; however, in such areas, the sequence of Foraminifera in the section provides a reliable basis both for age determination and for correlation from one local area to another. Subsurface contour maps are frequently drawn to show the first occurrence in the well of a diagnostic species of Foraminifera. These microfossil markers also serve to indicate the approximate vertical distance from such a datum to potential oil-producing horizons, which are known to produce petroleum in other fields. [D.J.J.]

Bibliography: M. W. Glaessner, *Principles of Micropaleontology*, 1947; Daniel J. Jones, *Introduction to Microfossils*, 1956.

Microphone

An electroacoustic device containing a transducer which is actuated by sound waves and delivers essentially equivalent electric waves.

Modern conventional microphones may be classified as pressure, gradient, combination pressure-gradient, and wave types. A pressure microphone is one in which the electrical response is caused by variations in pressure in the actuating sound wave. In a gradient microphone the electrical response corresponds to some function of the pressure difference between two points in a sound wave. A wave microphone is one in which the response depends upon sound wave interaction.

This article will discuss transducers used in microphones, some of the common types of microphones, and the characteristics of microphones.

MICROPHONE TRANSDUCERS

A transducer is a device which, when actuated by power from one system, will supply power to one or more other systems (see TRANSDUCER). For the conversion of the acoustical variations into the corresponding electrical variations, the transducers most commonly used in microphones are dynamic, magnetic, electrostatic, piezoelectric, electrostrictive, or carbon.

Dynamic transducer. This consists of a moving conductor located in a magnetic field (Fig. 1). The motion of the conductor leads to the induction of an electromotive force (emf) in the conductor, the magnitude of the emf being proportional to the conductor's velocity. The conductor may be in

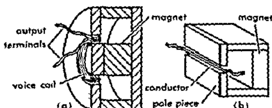


Fig. 1. Dynamic transducers. (a) Perspective sectional view of a moving coil transducer. (b) Perspective view of a straight line conductor.

the form of a coil, a straight wire, or a ribbon. The electrical impedance of the dynamic transducer is relatively low, ranging from 0.1 to 60 ohms in practical structures. The electrical impedance is practically all resistive and is therefore independent of frequency. When the dynamic transducer is connected to a load, it may be considered to consist of the open-circuit voltage, the impedance of the transducer, and the impedance ...

... is a variable reluctance path and a coil surrounding all or a part of this path. The variation in reluctance leads to a variation in the magnetic flux through the coil and a corresponding induced emf. There are an enormous number of arrangements and configurations possible using this principle. The two most common, namely, the single pole and armature and the multiple pole and armature types, are shown in Fig. 2. The electrical impedance of this type is governed by the size of the wire and the coil, and may vary from a few ohms to thousands of ohms. The impedance is proportional to the frequency, because the inductance of the coil is the predominant electrical element. When the magnetic transducer is connected to a load, it may be considered to consist of the open-circuit voltage, the impedance of the transducer, and the impedance of the load, all connected in series. The magnetic transducer is reversible.

Electrostatic transducer. The electrostatic or condenser transducer consists of a fixed electrode and a movable electrode. The electrodes are charged electrostatically in opposite polarity. Motion of the movable electrode leads to the production of a voltage which corresponds to the amplitudes of the electrode's motion (Fig. 3). The impedance of the electrostatic transducer is due to the capacitance between the two electrodes. When the electrostatic transducer is connected to a load, it may be considered to consist of the open-circuit voltage, the capacitance of the transducer, and the impedance of the load, all connected in series. The electrostatic transducer is reversible.

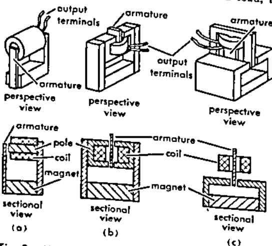


Fig. 2. Magnetic transducers. (a) Single pole and armature magnetic transducer. (b, c) Multiple pole and balanced armature magnetic transducer.

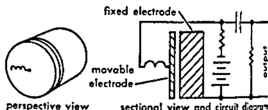


Fig. 3. Perspective and sectional views and electrical circuit diagram of an electrostatic transducer.

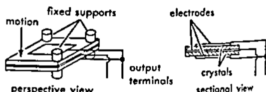


Fig. 4. Crystal transducer.

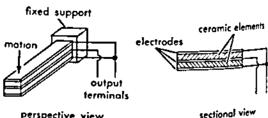


Fig. 5. Ceramic transducer.

Piezoelectric transducer. The piezoelectric or crystal transducer consists of a crystal element having piezoelectric properties (*see* PIEZOELECTRICITY). A deformation of the crystal leads to the generation of a voltage which corresponds to the amplitude of the deformation (Fig. 4). The impedance of this type is due to the electrical capacitance of the crystal. When the crystal transducer is connected to a load, it may be considered to consist of the open-circuit voltage, the capacitance of the crystal, and the electrical impedance of the load, all connected in series. The crystal transducer is reversible.

Electrostrictive transducer. The electrostrictive or ceramic transducer consists of a ceramic element having electrostrictive properties (*see* ELECTROSTRICTION). A deformation of the ceramic leads to the generation of a voltage corresponding to the amplitude of the deformation (Fig. 5). The impedance of this type is due to the capacitance of the ceramic. When the electrostrictive transducer is connected to a load, it may be considered to consist of the open-circuit voltage, the capacitance of the ceramic, and the electrical impedance of the load, all connected in series. The ceramic transducer is reversible.

Carbon transducer. This consists of carbon granules in contact with a fixed electrode and a movable

electrode (Fig. 6). Motion of the movable electrode varies the resistance of the granules. If the transducer is connected in series with an external steady voltage and a resistor, a voltage corresponding to the amplitude of the movable electrode's motion will be developed across the resistor. The impedance of this transducer is practically a pure electrical resistance, which is governed by the dimensions of the carbon granule aggregate. In general, the resistance is of the order of 100 ohms. The carbon transducer is not reversible.

MICROPHONE TYPES

Microphones may be classed in many different ways, as for example, by the type of electrical response, by the type of transducer, or by the directivity pattern (*see* DIRECTIVITY). In the subsequent discussion, microphones will be primarily classified according to the type of response.

Pressure microphone. This is a microphone in which the electrical response is caused by variations in sound pressure. Pressure microphones are inherently nondirectional (omnidirectional), because pressure is a scalar and not a vector quantity. Three of the most common directivity patterns of microphones are shown in Fig. 7.

Dynamic types. The three principal types of dynamic microphones are the moving coil, the moving conductor, and the ribbon-type pressure microphone.

A moving coil microphone consists of a voice coil located in a magnetic field and coupled to a diaphragm acted upon by sound waves (Fig 8). To obtain constant output for constant sound pressure on the diaphragm, the velocity of the coil must be independent of the sound wave frequency. To accomplish this objective, the system must be acoustical-resistance controlled. The resonant frequency

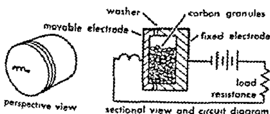


Fig. 6. Perspective and sectional views and electrical circuit diagram of a carbon transducer.

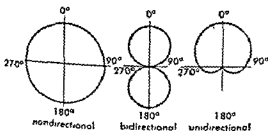


Fig. 7. Nondirectional, bidirectional, and unidirectional polar directivity patterns.

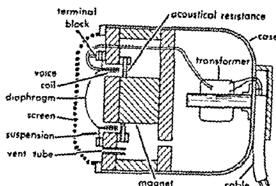


Fig 8 Sectional view of a moving coil microphone.

of the system is usually set at 700 cycles. The acoustical resistance is made larger in magnitude than the acoustical reactance due to the mass of the diaphragm and coil and the compliance of the suspension by placing an acoustical resistance behind the voice coil. This acoustical resistance may be a slit, silk cloth, or piece of felt.

The materials most commonly used for the diaphragms of pressure microphones are aluminum alloys, paper, Bakelite, styrol, and other plastics. In order to obtain a minimum density-resistivity product, aluminum is almost universally used for the voice coil The aluminum may be in the form of wound ribbon or wound wire.

The impedance of the voice coil may vary from 1 to 30 ohms. In general, a transformer is included to step up the impedance to any of the standard values used for transmission over a line, namely, 60, 150, and 250 ohms, or to a high impedance (25,000 ohms) for direct feed to the grid of a vacuum tube

For many applications an unobtrusive microphone is desirable In the unobtrusive design, a small dynamic unit is located in one end of a long slender tube (Fig. 9). The long tube provides sufficient volume to ensure adequate low-frequency response.

A small microphone supported by means of a slender band around the user's neck in the form of a pendant is known as a lavalier microphone. The miniature dynamic unit is housed in a case similar to that in Fig. 9, except that the diameter is smaller and the length is shorter. The principal purpose of the lavalier microphone is to allow a person to move freely without introducing any appreciable change in output, as would be the case if a stationary microphone were used. It also allows the talker to use his hands freely.

A moving conductor microphone consists of a straight-line conductor located in a magnetic field and coupled to a V-shaped diaphragm acted upon by sound waves. The action of this microphone is similar to that of the moving coil microphone.

A ribbon-type pressure microphone consists of a metallic ribbon located in a magnetic field and terminated in an acoustical resistance (Fig. 10). To obtain constant output for constant sound pres-

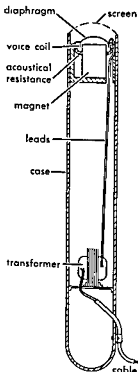


Fig 9. Sectional view of an unobtrusive moving coil microphone

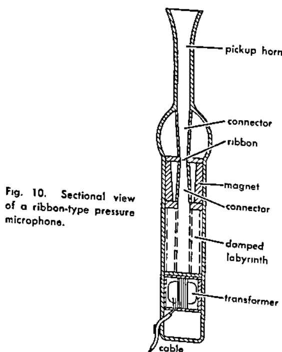


Fig. 10. Sectional view of a ribbon-type pressure microphone.

sure, the velocity of the ribbon must be independent of the frequency. This objective is accomplished by making the acoustical resistance, which terminates the ribbon, the controlling element. The acoustical resistance is in the form of a folded damped pipe. In order to improve the high-frequency response, a small horn is used at the pickup point to couple the ribbon to the sound field. A transformer is used to raise the electrical impedance of the ribbon to a value that is suitable for transmission over a line.

Magnetic type. A magnetic microphone consists of a diaphragm acted upon by sound waves and connected to an armature which varies the reluctance in a magnetic field surrounded by a coil (Fig. 11). To obtain constant output for constant sound pressure on the diaphragm, the velocity of the armature must be independent of the frequency Therefore, the system must be acoustical-resistance-controlled. This can be accomplished by placing an acoustical resistance behind the diaphragm In order to obtain stability of the magnetic armature system, considerable stiffness must be introduced into the system.

Electrostatic (condenser) type. An electrostatic or condenser microphone consists of a fixed plate and a movable plate or diaphragm exposed to the actuating sound waves (Fig. 12). The condenser (also called capacitor) microphone requires a high-voltage polarizing supply of the type shown in the figure. To obtain constant voltage output with respect to frequency, the amplitude of the diaphragm motion must be independent of the frequency. Therefore, the system must be stiffness-controlled. This can be accomplished by fixing the resonant frequency of the diaphragm at or above the upper frequency limit of the response range.

The capacitance of the condenser microphone is very small, being 50 micromicrofarads ($\mu\mu f$) to 300 $\mu\mu f$. The impedance is, therefore, very large at low frequencies. The polarizing supply resistors and the bias resistors must therefore be of the order of tens of megohms in order to maintain adequate low-frequency response. In view of this, it is almost imperative that the amplifier be located next to the transducer. In addition, a cathode follower type of amplifier can be used to provide a high impedance load for the transducer, as shown in the figure.

Crystal type. A crystal microphone consists of a crystal having piezoelectric properties acted upon by sound waves either directly or through a diaphragm connected to the crystal. A diaphragm-

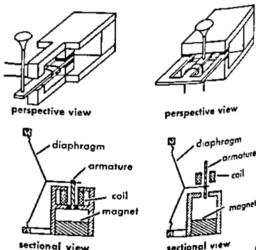


Fig. 11. Two magnetic microphones of the balanced-armature type. (*From H. F. Olson, Acoustical Engineering, Van Nostrand, 1957*)

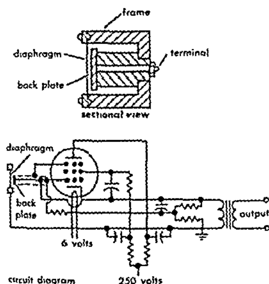


Fig. 12. Sectional view and vacuum tube electrical circuit diagram of a condenser microphone.

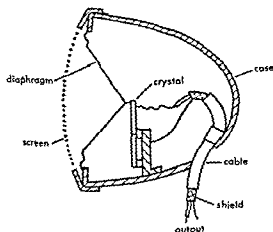


Fig 13. Sectional view of a crystal microphone.

type crystal microphone is shown in Fig. 13. To obtain constant voltage with respect to frequency for constant sound pressure, the amplitude of the crystal's motion must be independent of frequency. Therefore, the diaphragm must be stiffness-controlled. This is accomplished by placing the resonant frequency of the diaphragm and the crystal assembly at or near the upper limit of the response range.

The piezoelectric element may be Rochelle salt, or ammonium dihydrogen phosphate (ADP). In general, two crystals are cemented together so as to form a "bimorph" structure and thereby increase the ratio of stress to amplitude. In this way, the output is increased.

The capacitance of the crystal unit is of the order of 1000–2000 $\mu\mu f$. This makes it possible to transmit over several feet of low-capacitance cable without appreciable attenuation.

Ceramic type. A ceramic microphone consists of a ceramic, usually barium titanate, having electrostrictive properties. The ceramic is acted upon by sound waves transmitted through a diaphragm connected to the crystal. The general arrangement of the elements is similar to that of the crystal microphone shown in Fig. 13 except that a ceramic transducer is substituted for the crystal transducer. The performance and characteristics of the ceramic microphone are similar to those of the crystal microphone.

Carbon type. A carbon microphone consists of a diaphragm connected to a movable electrode which together with a stationary electrode forms a cup containing carbon granules (Fig. 14). A polarizing current is passed through the electrodes and the carbon granules. Motion of the diaphragm, caused by incident sound waves, varies the resistance of the carbon element. To obtain constant current variation in the electrical circuit of Fig. 14 for constant sound pressure upon the diaphragm, the amplitude of the diaphragm must be independent of the frequency. Therefore, the system must be stiffness-controlled. This can be accomplished by placing the resonant frequency of the diaphragm and carbon element assembly at or near the upper limit of the response range.

The electrical impedance of the carbon element is a resistance of the order of 100–200 ohms. A transformer is usually used with the carbon microphone, as shown in Fig. 14.

The carbon microphone is universally employed in all telephone applications. It is also used for communication systems where high sensitivity is a requirement. The second harmonic distortion is relatively high. Double button microphones, that is, microphones having two cups of carbon granules, have been used in a push-pull arrangement to reduce the nonlinear distortion.

Contact type. A contact microphone is one which can be attached directly to the body or sounding board of a musical instrument to provide solo pickup from the instrument. Contact microphones may employ magnetic, crystal, or ceramic transducers. The transducer is coupled directly to the instrument.

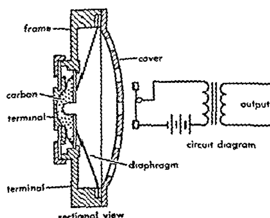


Fig. 14. Sectional view and electrical circuit of a carbon microphone.

Gradient microphone. A pressure-gradient microphone is one in which the electrical response corresponds to some function of the difference in pressure between two points in space. In general, when the distance between the two points is small compared to the wavelength, the pressure gradient corresponds to the particle velocity in the sound wave. A velocity microphone is one in which this condition holds.

Velocity type. A velocity microphone consists of a metallic ribbon which is located in a magnetic field and is freely accessible to sound waves on both sides. A schematic diagram of the elements of a velocity microphone is shown in Fig. 15. The ribbon is actuated by the difference in sound pressure between the two sides. The motion of the ribbon in the magnetic field induces a voltage between the two ends of the ribbon. If the effective distance between front and back is small compared to a wavelength, the magnitude of the pressure is the same on both sides of the ribbon. However, there is a phase shift between the two sides and hence a difference in pressure. The pressure difference is proportional to the frequency and also to the cosine of the angle of the incident sound with respect to an axis normal to the plane of the ribbon. In order to obtain constant voltage output, the velocity must be independent of frequency. Since the actuating pressure is proportional to frequency, the vibrating system must be mass-controlled to obtain a constant relationship between the sound pressure in free space and the velocity of the ribbon. This can be accomplished by placing the resonant frequency of the ribbon below the response range of the microphone. The resonant frequency is usually about 12 cycles in practical microphones.

One of the fundamental advantages of the ribbon-type velocity microphone is that the ribbon serves

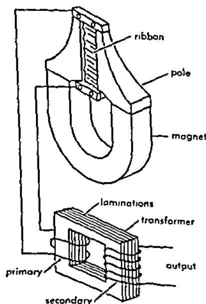


Fig. 15. Perspective view showing the elements of a velocity microphone.

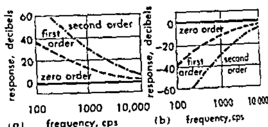


Fig. 16. (a) Frequency response of zero-, first-, and second-order gradient microphones at a distance of ¾ in. from a sound source. (b) Response characteristics of the same microphones for a plane wave. The microphones are compensated so that the responses of all three are the same and independent of the frequency when operating at a distance of ¾ in. from a sound source.

as both the diaphragm and conductor. For this reason, it is possible to obtain greater sensitivity than in a dynamic system, in which the diaphragm and voice coil have separate functions.

The directivity pattern of a velocity microphone is the bidirectional cosine characteristic shown in Fig. 7.

Higher order gradient types. The velocity microphone described in the preceding section is a ... order gra ... the ... tivity patterns are cosine functions in which the power of the cosine is the order of the gradient.

The response of a gradient microphone is a function of both the distance from the sound source and the frequency. The frequency response characteristics of a zero-order gradient or pressure microphone, a first-order or velocity microphone, and a second-order gradient microphone for a distance of ¾ in. are shown in Fig. 16*a*. Fig. 16*b* depicts the response of these microphones at a large distance from the sound source when the microphones are compensated to yield uniform response at ¾ in. This shows the discrimination of the gradient microphone against sounds originating at a distance. In addition, there is an additional 3 and 5 decibels attenuation by the first- and second-order gradient microphones, respectively, due to their directivity.

In first-order gradient microphones, both carbon and dynamic transducers have been employed to obtain discrimination against noise. A dynamic-type, first-order gradient antinoise microphone is shown in Fig. 17. Second-order gradient microphones use two first-order gradient microphones connected in opposition. A second-order gradient microphone using two dynamic-type units is shown in Fig. 18.

Unidirectional types. A unidirectional microphone is one having a substantially unidirectional pattern over its response range. Unidirectional microphones may be constructed by combining a bidirectional microphone and a nondirectional microphone or by combining a single-element microphone with an appropriate acoustical delay system

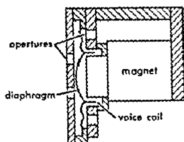


Fig. 17. Sectional view of a moving coil, first-order gradient microphone

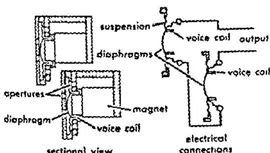


Fig. 18. Sectional view and electrical connections of a moving coil, second-order gradient microphone.

The directivity pattern of a unidirectional microphone is given by

$$e = a + b \cos \theta$$

where e is the voltage output of the microphone, a is the voltage output of the pressure element, b the voltage output of the velocity element, and θ the angle between the normal to the plane of the ribbon and the incident sound wave. If $a = b$ the directivity pattern is the cardioid pattern shown in Fig. 7.

Wave microphone. This consists of a system in which the directivity depends upon some type of wave interference. The most common wave microphones are line, reflector, and lens types. Line and reflector microphones are highly directional in the speech frequency range and may therefore be used for long-distance pickup of speech under conditions of high ambient noise and excessive reverberation. High-sensitivity transducers are employed in line and reflector microphones to provide an adequate signal-to-noise ratio.

Line type. A line microphone consists of a number of small tubes or pipes with the open ends, as pickup points, equally spaced along a line and the other ends connected at a common junction to a transducer. In the system shown in Fig. 19, the transducer is a ribbon element terminated in an acoustical resistance. Under these conditions, the output of the pipes can be added vectorially. In the case of a uniform line, the directivity characteristic, that is, the ratio R_θ of the response for an angle θ to the response for $\theta = 0$, is given by the equation

$$R_\theta = \frac{\sin \frac{\pi}{\lambda}(l - l \cos \theta)}{\frac{\pi}{\lambda}(l - l \cos \theta)}$$

where l is the length of line, λ the wavelength, and θ the angle between the direction of the incident sound and the axis of the line. The directional patterns for a simple line for line lengths of $\lambda/2$, 2λ, and 8λ are shown in Fig 19.

Reflector type. A reflector microphone consists of a surface which reflects the pencils of the impinging sound in phase to a common point, termed the focus, at which a microphone is located (Fig. 20). The directional characteristics of a parabolic reflector 3 ft in diameter are shown in the figure.

Hot-wire microphone. This consists of a fine wire heated by the passage of an electric current. The cooling due to the motion of air past the wire causes a change in electrical resistance of the wire. In a sound wave the particle velocity cools the wire. There are also minor cooling effects produced by the sound pressure The change in resistance due to the passage of a sound wave may be used to detect its presence. However, the frequency of the electrical output is twice the frequency of the sound wave because the wire is cooled equally by both positive and negative particle velocities. The use of a direct-current air stream for polarization appears to be impractical. Therefore, this microphone cannot be used for the reproduction of sound.

PERFORMANCE CHARACTERISTICS

The performance of a microphone is determined by the following principal factors: the open-circuit voltage response frequency characteristic, the electrical impedance characteristic, the directional characteristic, the nonlinear distortion characteristic, and the noise characteristic. The important characteristics of microphones are summarized in the table.

Open-circuit response. This is the open-circuit voltage output of a microphone as a function of sound frequency when the microphone is placed in

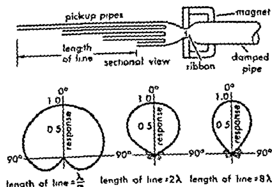


Fig. 19. Sectional view and directivity patterns of a line microphone. The maximum voltage response is arbitrarily chosen as unity.

Characteristics of microphones

Type	Directivity pattern	Transducer	Internal voltage	Internal impedance	Operating impedance*
Pressure	Nondirectional	Carbon	$e \propto x$	Resistive	L, M, and H
		Magnetic	$e \propto \dot{x}$	Inductive	L, M, and H
		Dynamic	$e \propto \dot{x}$	Resistive	L, M, and H
		Ribbon	$e \propto \dot{x}$	Resistive	L, M, and H
		Condenser	$e \propto x$	Capacitive	H
		Crystal	$e \propto x$	Capacitive	H
		Ceramic	$e \propto x$	Capacitive	H
Velocity	$\cos \theta$	Ribbon	$e \propto \dot{x}$	Resistive	L, M, and H
		Dynamic	$e \propto \dot{x}$	Resistive	L, M, and H
		Condenser	$e \propto x$	Capacitive	H
Second-order gradient	$(1 + \cos \theta)\cos \theta$	Ribbon	$e \propto x$	Resistive	L, M, and H
Antinoise	$\cos \theta$	Carbon	$e \propto x$	Resistive	L, M, and H
	$\cos^2 \theta$	Dynamic	$e \propto \dot{x}$	Resistive	L, M, and H
Line	Highly directional	Dynamic	$e \propto \dot{x}$	Resistive	L, M, and H
		Ribbon	$e \propto \dot{x}$	Resistive	L, M, and H
Reflector	Highly directional	Dynamic	$e \propto \dot{x}$	Resistive	L, M, and H

* Operating impedance is that obtainable with a transformer. L, low impedance, <100 ohms; M, medium impedance, >100 ohms and <1000 ohms; H, high impedance, >1000 ohms.

a plane wave sound field of constant pressure. It is expressed in decibels.

Electrical impedance. This characteristic is the complex ratio of the voltage applied across the electrical terminals to the resulting current as a function of frequency.

Directional characteristic. This is the open-circuit voltage output with respect to some reference axis of the microphone The directional characteristics may be depicted as either polar characteristics at different frequencies or frequency-response characteristics at different angles

Nonlinear distortion. This characteristic is the ratio of the total rms harmonic distortion output to the rms of the fundamental output. It may be expressed as a function of the frequency at different sound pressures or as a function of the sound pressure at different frequencies.

Noise characteristic. This is the open-circuit voltage generated in a given frequency band in the absence of an acoustical or electrical input. The noise spectrum level at a given frequency is the open-circuit voltage in a one-cycle bandwidth centered on that frequency in terms of the rms sound [illegible] the same voltage. [illegible] haracter [illegible] aracteristic and the transient response characteristic which are also important but more difficult to measure and specify. In some cases the type of cable used also affects the performance. [H F O]

Bibliography: L. L. Beranek, *Acoustics*, 1954. J. D. Moir, *High Quality Sound Reproduction*, 1958; H. F. Olson, *Acoustical Engineering*, 3d ed. 1957.

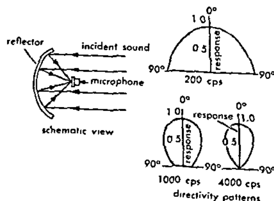


Fig. 20. Schematic view and directivity patterns of a parabolic reflector-type microphone. The directivity patterns are for a reflector 3 ft in diameter.

Microphonic tube

Electron tubes tend to be microphonic; that is, they have a tendency to respond unfavorably to mechanical vibration. This property is quite undesirable, particularly in audio-frequency applications, where any vibration may be heard in the output. It is even possible for such tubes to act as microphones, picking up and amplifying sound.

For the most part the microphonic properties of tubes are considered undesirable, and every effort is made to eliminate them. This is generally done by stiffening the supports to the electrodes and by providing internal damping so that vibrations do not persist. A number of special tube types have been devised with the expressed intention of producing a nonmicrophonic tube; these include the disk-seal tubes and, in particular, tubes in which special attention has been paid to the mechanical structure.

The microphonic properties of tubes can at times be used to advantage. Thus, these tubes can be used to detect sound or mechanical vibrations and to amplify them. It is also possible to adapt such tubes so that they can be used as pressure indicators or strain gages. *See* VACUUM TUBE. [R.R.S]

Microphonics

The transformation of the energy of mechanical vibration into an electrical signal. A microphone is microphonic in that it converts the mechanical energy of sound waves into an electrical signal.

The term microphonics is commonly applied to the production of an undesirable electrical output by an electronic device undergoing mechanical vibration. Often this is due to the excitation by vibration of mechanical resonances in the structure of a vacuum tube. The resulting motion of the elements in the vacuum tube changes the electrical characteristics of the tube and causes it to produce an electrical output. If this output is connected ultimately to a loudspeaker, a tone can be heard; this tone is termed microphonic noise. See MICROPHONIC TUBE.

In order to avoid microphonics, special consideration is given to the mechanical design of the elements in certain types of vacuum tubes. In addition, electronic devices designed for use in high-vibration environments, such as aircraft radios, are shock mounted to reduce the vibration transmitted to the equipment. See SHOCK ISOLATION. [D.N.K.]

Microradiography

The process of producing enlarged radiographic images on photographic film (see RADIOGRAPHY). Microradiography is applicable to small metallurgical, biological, and other specimens with gross structures not easily visible on conventional 1:1 radiographs. Historadiography is a term used for the microradiography of biological specimens, and hence is a branch of histology. Several techniques of producing enlarged images with beams of x-rays have been designated by the general term x-ray microscopy. Microradiography, then, is a type of x-ray microscopy, more specifically designated contact microradiography, in which the specimen is in contact with a fine-grained photographic emulsion, so that the absorption image is photographically enlarged or viewed with a microscope See HISTORADIOGRAPHY; MICROSCOPE, OPTICAL; MICROSCOPE, X-RAY.

It was recognized as early as 1913 that an enlarged radiographic image could be obtained in two ways: (1) by increasing the distance from specimen to plate to secure inherent enlargement of divergent beams and (2) by contact microradiography. The first of these techniques for a long time proved to be unsatisfactory because of the lack of sharpness and definition of the image caused by the fact that the x-ray beam was not a point source. This difficulty has been overcome by the technique of projection microradiography, also termed shadow microscopy.

In microradiography, it is convenient to be able to generate a wide range of essentially monochromatic beams, instead of relying upon a series of x ray tubes with different targets. It is possible to generate such beams at will by exciting secondary fluorescent radiation from a variety of elements with a single high-intensity source of primary rays, such as the Machlett AEG-50 tungsten target tube (see X-RAY TUBE). The experimental arrangement is shown in Fig. 1. Obviously the fluorescent beams are much lower in intensity than primary rays directly used, but exposure times are by no means prohibitive, and there is a large gain in homogeneity of the beam and hence in sharpness and resolution of the microradiographic image.

Contact microradiography. The technique of contact microradiography (CMR) consists of (1) preparation of a sample in the form of a small piece generally unmounted with a thickness of 0.1mm or less for metals and other dense materials, or considerably greater if necessary for biological specimens; (2) placement of the sample in close contact with a fine-grained photographic film in a simple camera or in a light-tight envelope; (3) exposure to a suitable x-ray beam; (4) enlargement of the developed image, which appears as a black spot. Thus the photographic image itself takes the place of the usual specimen for microscopic observation. Graininess may begin to interfere at magnifications of 300, although satisfactory results are often obtained up to 800.

This image is, of course, a two-dimensional registration of a three-dimensional absorbing specimen, or a superposition of images from various layers in the specimen, in contrast with the almost purely surface record provided by optical photomicrography for the highly polished and etched surfaces of metals, or by electron microscopy for surface replicas of thick specimens. However, transmission of light in the optical or electron microscope through transparent, and possibly stained, specimens, such as biological microtomed sections, is more closely similar. It is advantageous in specimens of essentially constant chemical composition but varying densities and thicknesses to use absorption staining by impregnation with heavily absorbing elements (lead, mercury compounds, silver iodide). This increases contrast, as in the case of the medical use of barium sulfate in radiographic delineation of the intestinal tract.

Projection microradiography. One of the two methods of enlargement long recognized for microradiography was to utilize inherent image enlarge-

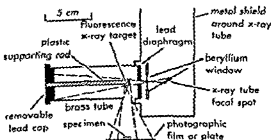


Fig. 1. Basic elements of microradiograph with selected wavelengths from fluorescent radiation.

ment (projection) by placing the film at various distances from the specimen instead of in intimate *contact with it*, thus requiring subsequent photographic enlargement of the image. But the projection image was always fuzzy because the x-ray beam was issuing from a focal spot of finite dimensions. The development of magnetic and electrostatic lenses to collimate electron beams into extremely fine pencils, as in the electron microscope, solved the problem since the electron beam impinging upon a target generated x-rays from what was essentially a point source. See ELECTRON OPTICS; MICROSCOPE, ELECTRON.

The successful development of projection microradiography (PMR) by V. E. Cosslett and W. C. Nixon followed in 1952, and then came commercial models of x-ray microscopes. Exceptional microradiographs of freeze-dried biological samples as well as of other materials have been widely published.

Applications. Since microradiography is a combination of radiography and micrography, analyses derived from it are at three levels: (1) for structure, which is most common and simplest; (2) for mass distribution; and (3) for mass-chemical analysis. The last two are accomplished best with thin biological specimens and with monochromatic beams of long wavelength.

Metallurgy and industry. There are many structural or morphological studies in which x-ray contrast mechanisms reveal detail not readily detected by optical microscopy. These include:

1. Metals of all types for micro-defects, gas porosity, microscopic cracks and cavities, segregations, cold-working.

2. Welds and coated metals of all types for complete soundness of junctions and bonds.

3. A[illegible] cially



Fig 2. Microradiograph of alloy containing Cu 9.40%; Si 0.73%; Fe 0.65%; Mn 0.030%; Mg 0.12%; Ni 0.10%; Zn 0.07%; Al remainder. Dark gray, primary Al dendrites; black, porosity; light, $CuAl_2$; white and most prominent, so-called Chinese-script phase α (Al, Fe, Mn, Si).

(Fig. 2). Complex alloys developed during and since World War II with as many as 10 components and 5 or more coexistent phases have been successfully characterized.

4. Storage battery plates and grids for porosity and bonding.

5. Electroplates for porosity, inclusions, and bonding.

6. Ceramic materials for uniformity, porosity, inclusions, incipient cracks, bonding, devitrification, foreign colloidal particles, and incipient failures in insulations.

7. Distorted crystals. If the incident x-ray beam is reflected by some properly oriented planes in accordance with Bragg's law (see X-RAY POWDER METHODS), there will be a distribution of light and dark areas on the microradiograph indicative of the existence and distribution of strain, which will change sensitively with tilting of the specimen.

8. Ores and minerals, especially those too opaque for examination with the (optical) polarizing microscope.

9. Formation of crystal nuclei in many problems involving earliest stages of crystallization.

10. Control of rubber manufacture: sulfur in vulcanization, carbon black, filler distribution, and particle size; behavior on stretching for incipient ruptures or strong adhesion between solid particles and rubber matrix; uniformity of blending of natural and synthetic rubber, whose absorption coefficients differ.

11. Identification of origin of sawdust and wood pieces, in comparison with standards for known materials (Fig. 3).

12. Plastic impregnation of paper, laminated wood, and so forth for complete penetration and bonding; sodium silicate bonds in corrugated fiberboard; flow lines and adherence of all types of adhesives.

13. Specifications and identifications of cigarette papers and filters, and the functioning of the latter in removing tars and resins and particulate matter.

14. Measurement of gage of foils and of density and uniformity of paints, varnishes, and other coatings.

15. Gage, uniformity, defects, and behavior of extremely fine fibers of glass and textiles.

16. Differentiation of papers, postage stamps, and paintings (originals vs. counterfeits).

17. Inspection of grains of barley, rice, corn and other materials used in brewing, and of seeds for insect infestation.

Biology (historadiography). The potentialities of microradiography in biological and medical research are obvious. They range from small insects and organisms, through tissues of all types—generally absorption-stained—both normal and pathological (cancerous), blood vessels (microangiography), circulation of fluids in animal and plant tissues, incipient bone pathology, lead poisoning, psoriasis, inflammatory destruction, aging, decalcification of bone, deposition of iron in cell nuclei in hematite miners, iodine content of thyroid folli



Fig. 3. Microradiographs ...

cles, incipient *disease* in plants such as Dutch elm disease, and any and all other diagnostic attempts for which magnification of the conventional radiograph yields *additional* information or simpler interpretation.

Special techniques. A number of special microradiographic techniques, applicable both to life science and industrial testing, have been developed. Some of these are described in the following paragraphs.

Stereoscopy. Just as in the case of radiography, two microradiographs may be made of a specimen in properly shifted position and then viewed in a stereoscope to gain a third-dimensional impression. A camera for this purpose was originally designed by the author and R. W. Eyler in 1943, and commercial equipment is now available.

Grainless media. The greatest single limitation on CMR is the graininess of the silver halide photographic emulsions, even in the Lippmann and similar types; an essentially grainless recording medium would permit much higher magnifications without interference with sharpness of detail. Two general types of grainless media are now proving to be successful. Upon irradiation with x-rays, polymers such as polyethylene and polyvinyl undergo a change in which the long polymer molecules are cross-linked; their solubility in appropriate solvents is decreased in proportion to exposure, while their tensile strength and electrical resistance is increased. Thus a latent image of a microradiographic specimen is formed on a layer of one of these plastics (the layer replacing a regular photographic emulsion); this latent image can be "developed" into a relief image by differential solubility in alcohol mixtures. This image can be replicated, and shadow cast and photographed in the electron microscope at thousands or hundreds of thousands diameter magnifications. Truly grainless photosensitive dyes are being successfully developed. Similarly, some crystal faces, such as those of ammonium dichromate, are radiation-sensitive and record laten... with proper solvents.

Some attempts also have been made to develop images on phosphors such as silver-activated metaphosphate glasses and Zn_2SiO_4 with 3% Mn activation; for high-resolution microradiographs with soft radiation, the photomicrographic exposures are excessively long. This tremendously important development is still in its early stages, but already the scope of microradiography has been greatly extended. H. H. Pattee has developed a very fine-grained fluorescent screen for the microfluoroscope, a related technique of greatest potential possibilities.

Color film. In the author's laboratory, unexpectedly promising results have been obtained for microradiographs on Ektachrome film. The color developed by x-rays (thus far all colors but red) may

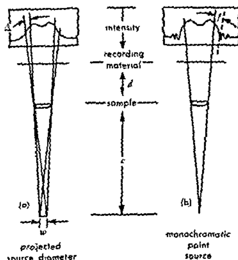


Fig. 4. Projection geometry as used in x-ray microscopy, illustrating two possible sources of resolution error. (a) Penumbra error. (b) Diffraction error.

be in terms of hue, value (brightness or reflectance), and chroma or saturation. These are dependent both on beam intensity and quality or wavelength; consequently, the developed microradiograph varies in hue, as well as value and chroma, which adds another dimension or method of differentiation.

Electronic microradiography. The emission of electrons from irradiated objects is used to produce images of very thin objects such as onionskin and butterfly wings, and even to copy photographs. In one technique hard (high-voltage) x-rays pass through the photographic emulsion without affecting it and strike the specimen; the back-reflected electrons whose energies and numbers correspond to the composition and texture of the specimen are easily registered by the film. In a second technique the x-ray beam liberates electrons from a thin lead screen, and these electrons are transmitted through the specimen to the photographic film. Electrons are far more easily absorbed than even soft x-rays and for some purposes in very thin specimens may give superior results. [C.L.CL.]

Instrumentation. The x-ray microscopes used in microradiography utilize x-radiation to form images of resolution in the 0.2–2.0 micron (μ) range and permit useful magnifications of 100–1000 diameters. Several types of x-ray microscopes have been developed, but only one type has been outstandingly successful—that which employs image projection (Fig. 4). With this method, x-ray microscopy has not only become competitive with direct light microscopy in resolving power but has gained important advantages: (1) New information can be obtained from unstained samples by virtue of the distinctly different x-ray contrast mechanism. The new technique of staining with radioactive stains is also introduced (2) The sample is replaced by a two-dimensional x-ray image, called a microradiogram, in which all levels of the sample object are sharply imaged into essentially a single focal plane for subsequent microphotographic enlargement. Such an optical "model" is often more amenable to high-resolution analysis, and it also allows precise stereographic measurement. (3) X-ray image contrast is simply and quantitatively related to the mass per unit area and to the mass absorption coefficient of the sample so as to permit micromass and microchemical analysis within areas as small as a few square microns.

Sources of error In order to resolve detail in the 0.2–2 0 μ region, not only must the penumbra error Δ and diffraction error f (Fig. 4) be sufficiently small, but also, for precise analysis, approximately 50% absorption signal must be obtained through a proper choice of the x-ray wavelengths. The optimum wavelengths for the x-ray microscope lie in the 5–50 angstrom (A) region. Fortunately, the critical absorption edges for elements of atomic numbers up to about 30 are also in this ultrasoft x-ray region. (For an explanation of absorption edges, see X-RAY FLUORESCENCE ANALYSIS.) For these wavelengths, the image spreading effects which limit resolution are essentially due only to the penumbra error and to Fresnel diffraction. The penumbra error is controlled by limiting the source size. For CMR, in which the sample lies in close contact with the recording material, the diffraction error f is negligible ($<0.1\ \mu$). For PMR, in which primary magnification (typically 40X) is obtained by placing the sample very near the source and far from the recording material, the diffraction error limits the resolution at these wavelengths, f being in the range 0.5–2.0 μ.

The greatest problem in the design of x-ray microscopes is the result of the extremely low production efficiency for ultrasoft x-radiations, which severely limits the image intensity. An important figure of merit, therefore, is the camera speed S, defined as the reciprocal of the time required to record a given sample, and given by

$$S = \frac{K i_o \Omega w^2 \Delta_r^2}{M^2 A} \qquad (1)$$

where K is a constant, i_o the radiant flux per unit solid angle and per unit projected source area; Ω the solid angle of the utilized x-radiation; w the source diameter; Δ_r the resolution error of the recording material; M the primary magnification; and A the sample area. For an optimum condition of operation, the penumbra error Δ is equated to the recording error Δ_r. Also, from Fig. 4 it is noted that

$$\Delta = wx = M\delta = (1 + x)\delta \qquad (2)$$

where $x = d/c$ and δ is the error Δ as measured at the sample. The camera speed may then be written as

$$S = K i_o \delta^4 (\Omega/A)(1 + 1/x)^2 \qquad (3)$$

Because of the dependence upon δ^4, it is important that the microscope and the recording material be adjustable so as to present no higher resolving power than that required for the problem under investigation. Also, it is important to note the rapid increase of efficiency for the CMR region.

Microscope types. There are three basic designs for projection microscopes:

1. CMR for large areas. By making the source-to-sample distance large (10–20 *cm*) in order to cover a large sample area, the time required for the total exposure becomes proportionately long. It is then necessary to use photographic recording material (concentrated Lippman emulsions) because of its relatively high sensitivity. The large working distance permits the use of a rotating sector to provide a calibration wedge for quantitative analysis (Fig. 5 shows an instrument that exposes 2- by 3-in. plates).

2. CMR for small areas. A considerable gain in camera speed is effected by placing the sample and recording material a small fraction of a millimeter from a microfocus source. In microfocus tubes i_o is limited by the maximum emission current of the electron source, and can be 5–50 times that for

the large focal spot sources. With this arrangement, the recording material might also be a thin, fine-grained phosphor (such as Mn-activated Zn_2SiO_4) evaporated on a cover slip so as to permit direct viewing. The image may be microphotographed or microphotometered directly from the fluorescent screen.

3. PMR. By placing the sample very near the microfocal spot and several centimeters from the recording material, primary magnification may be gained, but at a great loss in camera speed. Since direct viewing from a fluorescent screen is required for selection of field and for focusing the electron beam, the necessary x-ray intensity is gained only

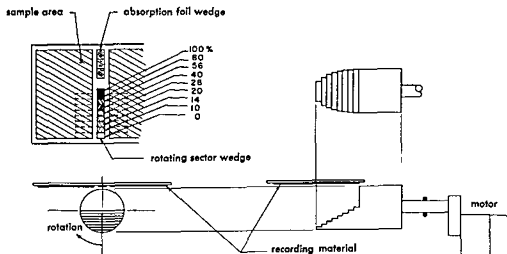


Fig 5. Ultrasoft x-ray source and camera for microradiographic analysis. A, high vacuum isolation gate; B, demountable anode; C, grid and focusing electrode; D, helical tungsten cathode; E, charcoal trap. *(After B L. Henke, High Resolution Microradiography, AFOSR Tech. Rept. 2, 1958)*

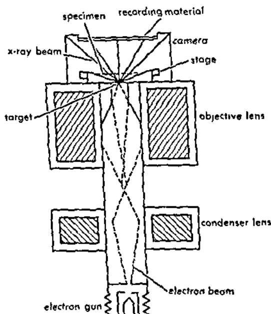


Fig. 6. PMR microscope. (After W. C. Nixon, V. E. Cosslett, A. Engström, and H. H. Pattee, eds., X-Ray Microscopy and Microradiography, Academic Press, 1957)

by using wavelengths less than 10 A. This may permit a reduction in the diffraction error, but it also limits application to samples of high atomic number or to thick samples which require relatively low resolving power (for instance, living macroorganisms). Generally, with PMR and using conventional x-radiation, it is not required to subject the sample to reduced pressures or vacuum. An example of a PMR instrument is shown in Fig. 6.

[B.HE.]

Bibliography: G. L. Clark, *Applied X-rays*, 4th ed., 1955; V. E. Cosslett, A. Engström, and H. H. Pattee [illegible]
raphy,
X-ray [illegible]
lution [illegible]
1958; B. L. Henke, R. White, and B. Lundberg, *J. Appl. Phy.*, 28(1):98–105, 1957; H. H. Pattee, *Science*, 128(3330):977–981, 1958; J. Trillat, *Exploring the Structure of Matter*, 1959.

Microsauria

An order of Carboniferous and early Permian amphibians with spool-shaped lepospondylous vertebrae and a body build of essentially normal type, but with a tendency toward elongation and limb reduction. Microsaurs were generally of small size. They have sometimes been suggested as possible ancestors of all or part of the class Reptilia, but such features as a reduced pattern of skull elements and absence of intercentra in the vertebrae render this improbable. On the other hand, the microsaurs may well be ancestral to the modern orders Urodela (Caudata) and Apoda (Gymnophiona). See AMPHIBIA FOSSILS; LEPOSPONDYLI; see also GYMNOPHIONA. [A.S.R.]



Microbrachis, a Late Pennsylvanian microsaur. (After Steen)

Microscope

An instrument used to obtain an enlarged image of a small object. The image may be seen, photographed, or sensed by photocells or other receivers, depending upon the nature of the image and the use to be made of the information of the image. [illegible] classified as simple or compound, [illegible] rm the image, [illegible] signed.

Simple microscope. A simple microscope, hand lens, or magnifier usually is a round piece of transparent material, ground thinner at the edge than at the center, which can form an enlarged image of a small object. Commonly, simple microscopes are double convex or planoconvex lenses, or systems of lenses acting together to form the image (Fig 1). The lens can be mounted in a simple holder, in a folding case for hand use, or with a support which has a mechanical focusing mechanism, stage, and mirror to make a dissecting microscope.

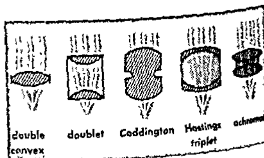


Fig. 1. Common types of magnifier. (From F. A. Jenkins and H. E. White, Fundamentals of Optics, 3d ed., McGraw-Hill, 1957)

Compound microscope. The compound microscope utilizes two lenses or lens systems. One lens system forms an enlarged image of the object and the second magnifies the image formed by the first. The total magnification is then the product of the magnifications of both lens systems. Theoretically, several simple microscopes could be used in line, each to magnify the image of the one before it. Practically, the losses from aberrations, reflections, and other defects limit the compound microscope to two such systems (Fig. 2).

The typical compound microscope consists of a stand, a stage to hold the specimen, a movable body-tube containing the two lens systems, and mechanical controls for easy movement of the body and the specimen. The lens system nearest the specimen is called the objective; the one nearest the eye is called the eyepiece or ocular. A mirror is placed under the stage to reflect light into the

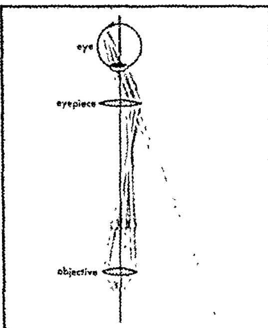


Fig. 2. Compound microscope diagram. (From F. A. Jenkins and H. E. White, Fundamentals of Optics, 3d ed., McGraw-Hill, 1957)

instrument when the illumination is not built into the stand. For objectives of higher numerical aperture than 0.4, a condenser is provided under the stage to increase the illumination of the specimen. Various optical and mechanical attachments may be added to facilitate the analysis of the information in the doubly enlarged image.

Special compound microscopes have two image-forming systems to give an enlarged image of an image. These instruments utilize electrons, x-rays, sound, or other forms of radiation for image formation and electromagnetic or electrostatic fields or mirrors to form the enlarged images. Because these images are not visible, photography, television, and special receivers must be used to record and analyze the image. The special microscopes are usually named according to the kind of radiation used, such as *electron, x-ray, ion,* and *ultrasonic.*

Types of compound microscope. The compound microscopes which employ a lens system are essentially similar in their working principles. Basically, they are modifications of the ordinary laboratory microscope (bright-field microscope) for specialized or specific purposes. The following is a list of some of the different types.

Light or photon microscopes utilize light of wavelengths from 380 to 760 mμ for image formation. Such microscopes include the laboratory or bright-field microscope, and modifications of it such as the capillary, centrifuge, chemical, comparison, crystallographic, dark-field, dissecting, fluorescence, integrating, interference, inverted microprojection, museum, nuclear track, petrographic, phase, phosphorescence, and profile microscopes.

Reflecting microscopes utilize a mirror rather than a lens system. The infrared microscope uses radiation of wavelengths greater than 700 mμ and the ultraviolet employs light of 180–400 mμ. The ultraviolet microscope requires reflecting optics or special quartz and crystal objectives. Buerger and color translating microscopes employ two and three different wavelengths of light, respectively, in the examination of specimens.

In the electron, proton, x-ray, and β-ray microscopes the image is usually recorded on a fluorescent screen or is photographed.

Mechanical vibrations, generated into an elastic system, provide the basis for the ultrasonic microscope employed for locating foreign bodies or the analysis of reflecting surfaces. *See* MICROSCOPE, CENTRIFUGE; MICROSCOPE, ELECTRON; MICROSCOPE, FLUORESCENCE; MICROSCOPE, INTERFERENCE; MICROSCOPE, OPTICAL; MICROSCOPE, PHASE CONTRAST; MICROSCOPE, REFLECTING; MICROSCOPE, X-RAY.

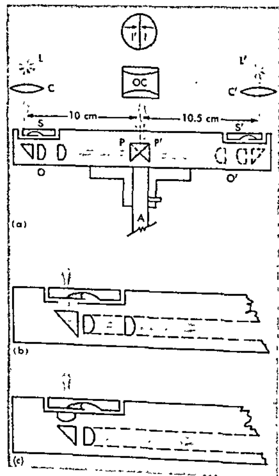
[O.W.R.]

Microscope, centrifuge

An instrument developed by E. N. Harvey and A. L. Loomis in 1930, which makes possible the magnification and observation of living cells or small organisms while they are being centrifuged. The fundamental optical principle involved is a mirror or prism system. The image of the material magnified by the objective, which rotates near the periphery of the centrifuge head, is brought to the axis of rotation where it is observed with a stationary ocular (see illustration *a*). In this respect it differs radically from the optical method used to observe or photograph the boundary of sedimenting protein molecules in an ultracentrifuge, where only a uniform distribution of material in one dimension (that along a radius of rotation) can be photographed. The combination of a centrifuge with the appropriate parts of a microscope provides for a perfect two-dimensional image of the material under observation. High dry objectives (43×) can be used and clarity of the image is independent of speed of rotation. The instrument can be used to observe stretching of living cells and movement of granules and other cell structures in a centrifugal field, thereby giving information on viscosity and surface forces of the cell.

It is not necessary to use stroboscopic illumination with the instrument. If the image of an incandescent lamp filament parallel to a radius of rotation is focused on the object to be examined, the image will be sufficiently clear over most of the microscope field. The reason is that a distance or a movement at the periphery becomes reduced near the center of rotation in the ratio: radius of centrifuge/radius in microscope field. Thus a movement of 100 μ under the illumination at the periphery may become less than 1 μ near the center. Flicker disappears at the higher speeds.

High- and low-power objectives can be mounted in the rotating head as shown in the illustration, *b* and *c*. An optical system for observation of con-



(a) Diagram of the optical system of a centrifuge microscope head for simultaneous observation of control and experimental living cells. *A*, axis of rotation; *L*, *L'*, incandescent filament light source; *C*, *C'*, condensing lenses to throw image of filament on special slides *S'*, *S*, containing the cells; *O'*, *O*, objective lenses and reflecting prisms revolving with centrifuge head; *P*, *P'*, central reflecting prisms; *OC*, stationary ocular; *I'*, *I*, microscope fields from the two ends of centrifuge head. (b) Arrangement of special microscope slide, lenses, and prisms in low-power objectives and (c) in high-power objectives.

and experimental material centrifuged simultaneously at opposite ends of the rotating head has been described by Harvey. Two sources of illumination and slightly different radii of rotation for the material to be examined are necessary.

A simplification may be made, for use with the Beams turbine, where it is not necessary to rotate objective lenses, but only the *mirrors*. The image of a straight-filament incandescent lamp is focu[illegible] on the materi[illegible]

[illegible]

[illegible] the organisms in position in approximately the focus of the objective. Final focusing is carried out by moving the ocular or an accessory lens between the ocular and central prisms. [E.N.H.]

Bibliography: E. N. Harvey, A centrifuge-microscope for super-centrifugal forces, *Science*, 75:267, 1932; E. N. Harvey, The microscope-centrifuge and some of its applications, *J. Franklin Inst.*, 214:1–23, 1932; E. N. Harvey, A new form of centrifuge-microscope for simultaneous observations of control and experimental material, *Science*, 77:430, 1933.

Microscope, electron

A device for forming greatly magnified images of objects by means of electrons. Electron microscopes serve two purposes: (1) the visual examination of structures too fine to be resolved with ordinary, or light, microscopes, and (2) the study of surfaces that emit electrons. The first function has made transmission electron microscopes, such as that shown in Fig. 1, essential research tools in biology, chemistry, and metallurgy.

INSTRUMENTS

A transmission electron microscope, shown in close analogy with a light microscope in Fig. 2, consists in its simplest form of a source supplying a beam of electrons of uniform velocity, a condenser lens for concentrating the electrons on the specimen, a specimen stage for displacing the specimen which transmits the electron beam, an objective lens, a projector lens, and a fluorescent screen on which the final image is observed. For permanent record of the image, the fluorescent screen is replaced by a photographic plate or film.

Electrons are strongly scattered by all forms of matter. Hence the entire instrument must be evacuated to about 10^{-4} mm Hg (10^{-7} atmospheric pressure). Furthermore, the lenses cannot be material in nature. Instead, they are electric or magnetic fields, symmetrical about the axis of the instrument, which have the property of bending the electron paths toward the axis, just as converging glass lenses bend light rays toward their axis. The lenses in the instruments shown in Figs. 1 and 2 are magnetic fields formed at narrow gaps in the iron casings surrounding coils traversed by electric current. The lens strength is varied by varying the coil current. Most electron microscopes employ magnetic lenses of this type. These have yielded the highest resolution and magnification attained so far.

However, excellent results have also been obtained with electron microscopes employing unipotential electrostatic lenses and magnetic lenses excited by permanent magnets. *See* ELECTRON LENS; ELECTROSTATIC LENS; MAGNETIC LENS.

Resolution. A microscope can, at best, permit the discrimination of two point objects greater than $0.6\lambda/\sin\theta$ apart. Here λ is the wavelength of the

illuminating radiation and θ is the aperture angle of the cone of radiation that participates in forming the image (Fig. 3). For green light $\lambda = 5000$ angstrom (A). One angstrom equals 10^{-8} cm. Even for ultraviolet radiation in an immersion medium of refractive index 1.5, λ is no less than 1700 A.

Since light and ultraviolet microscope objectives can be designed to utilize practically all the radiation passing through the specimen, $\sin \theta \approx 1$ and the least resolvable distance for the ultraviolet microscope is about 1000 A. For 50-kilovolt electrons, such as are commonly employed in electron microscopes, the wavelength is only 0.05 A. Hence, even though a cone of radiation with an aperture angle less than 0.01 radian contributes to an image of optimum sharpness, object separations smaller than 10 A have been resolved with the electron microscope. Thus the electron microscope has over a hundred times the resolving power of the light microscope. Similarly, whereas the maximum useful magnification of the light microscope is about

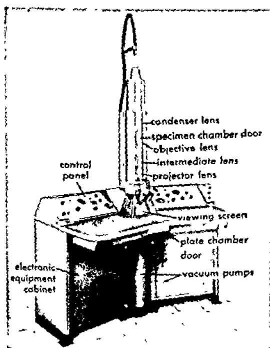


Fig. 1. Electron microscope. (*Radio Corporation of America*)

2000, that of the electron microscope is about 200,000. The maximum useful magnification is the least magnification of the image that reveals to the observer all the specimen detail which the microscope is capable of conveying.

Transmission electron microscope. The electrons are commonly emitted from the tip of a fine tungsten-wire hairpin filament, which is maintained at a carefully stabilized negative potential of 30–100 kv with respect to the remainder of the instrument. The electrons enter the instrument through an anode aperture. The intensity and convergence of the electron beam falling on the specimen are adjusted by varying the coil current of the condenser lens.

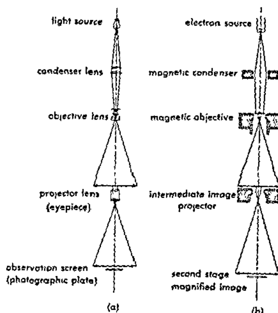
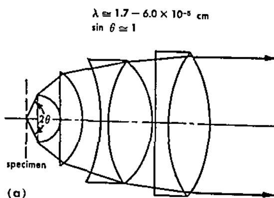


Fig. 2 Comparison of (a) light microscope and (b) magnetic electron microscope. (*From V. K. Zworykin et al., Electron Optics and the Electron Microscope, Wiley, New York, 1945*)

Image contrasts are formed by the scattering of electrons out of the narrow cone that contributes to the formation of the image; denser or thicker portions of the specimen scatter more electrons and hence appear darker in the image. The sharpness of the image observed on the screen is adjusted by varying the objective coil current, its magnification by varying the projector coil current. Both currents must be carefully stabilized to yield high resolution. In most modern instruments the function of the projector is performed by two magnetic lenses in tandem. In the instrument shown in Fig. 1 a magnification range of 20:1 is obtained by varying the coil currents only; by the exchange of pole pieces, a total magnification range from 700 to 200,000 may be covered. Shortened exposures and a larger field of view are obtained by recording the image on fine-grain plates at a lower magnification and enlarging the negatives. A low-power observing microscope aids in adjusting the objective current for maximum image sharpness.

Transmission electron microscopes that deviate from the standard form just described are the shadow electron microscope and the scanning electron microscope (Fig. 4). In both these instruments, which have achieved only limited practical usefulness, electron lenses are employed



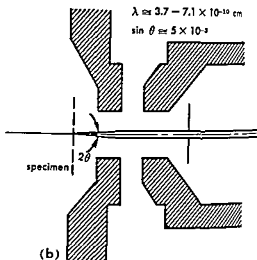
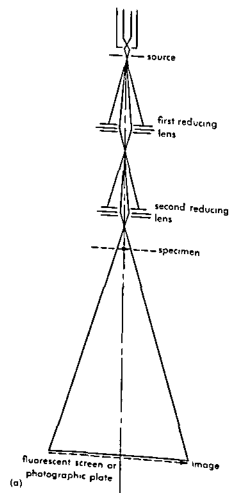


Fig 3. Effective objective aperture in (a) light microscope and (b) electron microscope.

greatly reduced images of the electron source. In the shadow electron microscope, the specimen is placed a small distance in front of the source image, which projects an enlarged shadow image of the specimen on a screen or photographic plate.

In the scanning microscope, the source is imaged directly onto the specimen and is displaced across it in a minute scanning pattern by a deflection field applied to the beam. The transmitted electrons (or secondary electrons ejected by the beam from



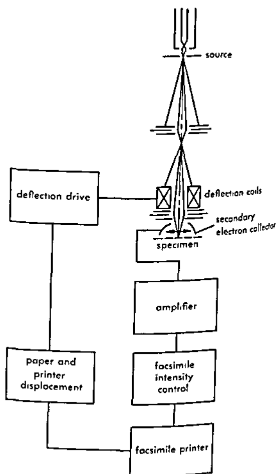


Fig. 4 Schematic representation of (a) shadow microscope and (b) scanning microscope.

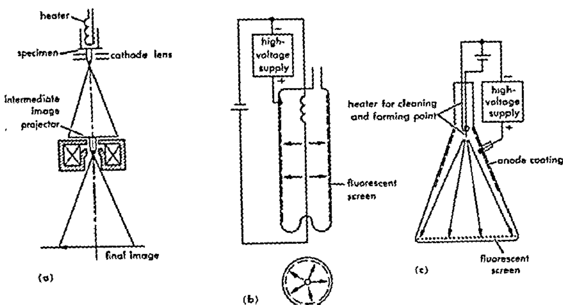


Fig 5. Emission electron microscopes: (a) immersion electron microscope, (b) cylindrical thermionic-emission microscope, (c) field-emission microscope.

the specimen) modulate the intensity on a much larger facsimile pattern traced in synchronism with the pattern on the specimen. See FACSIMILE.

Emission electron microscopes. These also take several forms (Fig. 5). In the immersion electron microscope the specimen is a flat metal surface, which may be heated, illuminated, or bombarded by high-velocity electrons or ions so as to emit low-velocity thermionic, photo-, or secondary electrons. These are accelerated to a high velocity in an *immersion objective* or cathode lens and imaged as in a transmission electron microscope.

Other emission microscopes employ simple projection without lenses. One example is the cylindrical microscope used for examining the thermionic emission from a thin wire stretched along the axis of a cylindrical tube whose walls, at high positive potential, are coated with a fluorescent substance. Another is the field-emission microscope, in which field electron currents are drawn from an extremely fine, rounded point on the end of a wire of a metal such as tungsten. The electrons drawn out of the point follow nearly straight lines toward a positive-potential fluorescent screen. The display on the screen shows variations in the emission characteristics (as determined by the work function and surface nonuniformities) over the surface of the point. See FIELD-EMISSION MICROSCOPY; THERMIONIC EMISSION. [E.G.RA.]

ELECTRON MICROSCOPY

The very different character of the interaction of electrons and light with matter produces the greatest practical differences in electron microscopy and optical microscopy. Contrast in the optical microscope image is produced largely by spectral absorption of the light passing through the specimen. Contrast in the electron microscope image is produced primarily by the scattering of electrons from the specimen. Electrons in the 50–100-kv range used in electron microscopes are strongly scattered by matter, and it is this strong interaction which places serious limitations on specimen preparation. As indicated previously to avoid scattering of the electrons by the ambient gas molecules, the electron beam must be confined to a column evacuated to a pressure of the order of 10^{-4}–10^{-5} mm of Hg. Thus all specimens must be prepared in such a manner that they can be placed in a good vacuum. Complete dehydration is required and this requirement immediately places a severe restraint on all biological systems. No living organisms have been observed in the electron microscope. Secondly, because of the strong scattering of the electrons by matter, the total amount of material that can be allowed in a specimen is very small. The specimen must be extremely thin. Substrates for mounting specimens are typically of the order of 100 A thick. To examine tissue, sections must be cut which are no thicker than 0.1 micron (μ). Specimens containing too much matter cannot be examined in the electron microscope. As the amount of material in the specimen increases, multiple scattering destroys the contrast in the image and finally with too much matter the energy transferred from the beam by inelastic collisions becomes prohibitively large and the specimen is destroyed by the electrons irradiating it.

Another important consequence in electron microscopy which follows from the nature of the electron interaction with matter is the limitation on selective staining of specimens. The effective use

of colored dyes in optical microscopy selectively to introduce contrast in given chemical structures is well known. In electron microscopy a monoenergetic beam must be used because of the strong chromatic aberrations of the lenses, and contrast obtained depends only on the differences in thickness and density in the specimen, rather than on differences in chemical composition. Selective staining must depend upon introducing variations in density, a method which is limited compared to the rich range of colored dyes available to the optical microscopist.

Specimen preparation. The following sections consider various aspects of specimen preparation for electron microscopy.

Substrates. The material to be examined in the electron microscope must, in general, be mounted on an extremely thin supporting membrane which has a mass density per unit area much less than that of the object being observed. In practice a membrane of the order of 100 A thick is mounted on a fine electrodeposited metal mesh with openings 50–100 μ on a side and the specimen material is deposited on these mounted substrates. Cellulose nitrate films prepared by casting a film on water from a solution of cellulose nitrate in amyl acetate produces a suitable substrate. Another substrate material is "Formvar," a polyvinyl formal plastic, cast on glass from solutions of the plastic in dichloroethane (ethylene dichloride). More durable substrates can be prepared by evaporating carbon on a glass slide, stripping off the carbon film by floating it onto a water surface and mounting the film on the standard mesh specimen screens. Carbon films substrates as thin as 50 A can be produced and handled.

Shadowcasting. Materials of low mass density will not produce appreciably greater scattering than the substrate film upon which they must be mounted. Contrast in electron images from these specimens will therefore be low, and since this low differential of mass density applies especially to organic materials, it presents a large obstacle in imaging very small biological objects. The technique of shadowcasting developed by R. C. Williams and R. W. G. Wyckoff is a very powerful means of enhancing the contrast in the electron microscopy of small objects of low mass density. The technique has been developed and applied to the point where it is now possible to resolve macromolecular material down to the 20–30-A range.

The technique involves the vacuum deposi-

receives a heavier deposit of the metal on the side facing the metal evaporating source. The metal shadowing layer produces a topographical representation of the surface of the specimen, and, as discussed below, is now an essential step in all replication of surfaces for study by electron microscopy (Fig. 6).

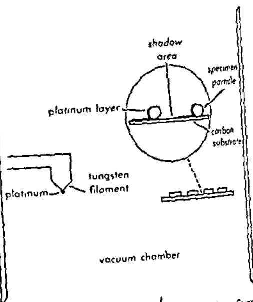


Fig. 6. Schematic of shadowcasting procedure. Source to specimen distance is commonly 10–20 cm; shadow angle, 1:4.

After shadowcasting, the metal layer is the specimen imaged by the electron microscope, for the scattering from the shadowcast specimen is predominantly that of the deposited metal layer. The areas in the shadow where no metal is deposited will scatter relatively few electrons and the respective areas of the image will appear dark on the photographic plate which has been exposed to the high electron density incident in these portions of the image. Projecting areas facing the evaporation source will have heavy layers of the metal and will scatter most electrons, thus appearing light on the photographic record of the image. An example of a shadowcast specimen is shown in Fig. 7, in which the image is reproduced in the customary manner with the same relative light and dark as on the original electron micrograph negative. The effect is as though strong illumination were falling obliquely on the specimen.

The metals most generally used are uranium, platinum, or chromium. The procedures developed for producing thin films by metal evaporation are applied. A tungsten filament, charged with the metal to be evaporated, is heated in a vacuum of about 1×10^{-5} mm Hg to a temperature at which the metal has a vapor pressure of the order of 1×10^{-2} mm Hg and a layer of mean thickness of about 5 A is deposited on the specimen at the desired angle.

Particulate materials. The examination of particulate material with the electron microscope requires the proper dispersion of a very small sample of the material on the substrate. Methods must be utilized which prevent aggregation and surface tension effects of drying. If the material to be examined is a dry powder, a dispersion technique must be employed. One method used extensively is to mill the powder in a viscose solution of cellulos

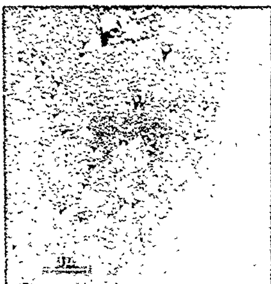


Fig. 7. Edge of spray droplet containing bacteriophage. For assay work the whole drop must be imaged.

nitrate in amyl acetate until the aggregates have been properly dispersed. The suspension is then diluted with amyl acetate to the proper concentration for casting a suitably thin film with the particles suspended in it.

Particles already dispersed and suspended in an aqueous medium can be deposited on plastic or carbon substrate films. The effects of surface tension on drying, the aggregation of the particles, and (with biological materials) the effects of changing salt concentration and pH on drying must be considered. A method of minimizing these effects is to spray the particle suspension on the substrate in very small droplets of the order of 5–10 μ in diameter. If buffers are needed, volatile ones such as ammonium acetate, ammonium carbonate, or veronal acetate must be used. A further advantage of the spray technique is the rapid drying of the small droplets, minimizing the effects of changing concentration on the biological materials. By depositing the droplets on substrates cooled to liquid nitrogen temperatures and subliming the ice under vacuum, the surface tension effects of drying can be further reduced.

The spray technique can be used for quantitative work with the electron microscope and assays can be made on suspensions of particles of unknown concentration in which a direct count is obtained on the number of particles per unit volume. Since each droplet is an aliquot sample of the total specimen, when the volume of the droplet is known, a count of the particles in a drop pattern gives the assay. By mixing the suspension to be assayed with a given volume of a known concentration of monodispersed polystyrene latex particles and counting the number of polystyrene in each drop, the volume of the drop is obtained. When a highly purified material is available for assay, it is possible to take an aliquot portion to dryness, weigh it, and from the electron microscope particle-count calculate the molecular weight of the particle. Figure 8 shows tobacco mosaic virus (rods) being assayed using the polystyrene spheres.

The electron microscope is capable of resolving macromolecules of biological interest. Spraying a purified sample of macromolecular material onto a very smooth surface such as freshly cleaved mica produces a suitably dispersed specimen. This specimen is shadowed directly on the mica with platinum, backed with an evaporated substrate of carbon and stripped off the mica using methods described below in the description of replica techniques.

Tissues and cells. The first successful thin sectioning using a modified conventional microtome was reported in 1948 by D. C. Pease and R. F. Baker. Since then the techniques have developed rapidly and the methods for study of the ultrastructure of cells and tissues have become highly developed (Figs. 9 and 10). Microtomes, using fractured glass or polished diamonds for knives, could reproducibly cut serial sections down to thicknesses as low as a few hundred angstroms. Techniques for fixation, dehydration, and embedding of tissue, though still in a rapid stage of development, are adequate for a large range of observations on the complex detail of cell structure. The most widely used fixative is buffered osmium tetraoxide, while others such as potassium permanganate have been used successfully for examination of certain cellular components. Dehydration is carried out through a series of alcohols of increasing concentration. The tissues are embedded in a plastic matrix by impregnating with a monomer which contains a suitable catalyst and polymerizing the monomer by heating or by exposure to ultraviolet radiation. Butyl methacrylate is widely used for embedding, while epoxy resins show some advantages when used for embedding certain tissues. Staining of tissue is being actively investigated. Though specific electron stains cannot be expected to produce the strong contrast effects available with colored dyes in optical microscopy, the literature contains an increasingly large number of contributions on electron staining of tissue. Here again the instrumental resolving power of the electron microscope exceeds the resolution of detail available in the specimen. Such fine structure as membranes with an apparent thickness of 25 A are observed, but the significance of such dimensions is still questionable considering the artifacts that may be introduced in the fixation, staining, and embedding procedures. In this rapidly developing field, the current literature must be consulted for an accurate evaluation of the present stage of achievement.

Surfaces by replication. Surfaces of materials too thick to be placed directly in the electron microscope can be examined in the transmission electron microscope by preparing a thin film replica of



Fig. 8. Spray droplet pattern for assaying tobacco mosaic virus (rods) using polystyrene latex particles (spheres).

the surface. The shadowcasting technique which produces contrast effects characteristic of the topography of the surface on which it is deposited greatly enhances the results obtainable by replication. Two of the most generally used materials for thin film replicas are cellulose nitrate and Formvar. Evaporated carbon replicas are to be preferred for their greater strength [illegible]

plo [illegible]
Eac [illegible]
fro [illegible]
hav [illegible] application to different surfaces. A relatively smooth surface may be replicated directly by coating a film of Formvar on it by flooding the surface with a 0.5% solution of Formvar in ethylene dichloride. For greater ease of stripping, the replica may be backed by coating a cellulose nitrate film on the Formvar film. The backed replica can be wet stripped, either by allowing the plastic film to float off on water if it will easily separate from the surface being replicated, or where necessary the material being replicated can be dissolved away with a suitable solvent or acid and the replicating film released. The replica film is mounted on specimen grids with the replication surface up, shadowed, and the backing film, if present, dissolved away in a solvent that does not attack the replica film. In the case of a Formvar replica backed with cellulose nitrate, amyl acetate is satisfactory. Dry stripping is accomplished by sandwiching the specimen grids between the film and Scotch tape and pulling off the film. Sometimes a stripping layer of a hydrophilic material is



Fig. 9. An electron micrograph of a longitudinal section of the spider's leg just above the tarsal-metatarsal joint cut through the vibration receptor (ear) of the spider. (M. Salpeter)

coated on the surface to be replicated. A detergent material such as Victawet can be evaporated onto the surface to be replicated, the metal shadowing deposited directly over the Victawet coat in the same vacuum chamber, and, to complete the replica, a carbon substrate deposited by evaporation. This preshadowed replica is perhaps the most satisfactory and is used if the material to be replicated allows. On very rough surfaces, such as wood and fractured materials, it may be necessary to form a thick negative replica of a material such as polystyrene which can be molded to the surface by heat and pressure. This negative replica can then be replicated by shadowing, evaporating a carbon film on it, and the replica released by dissolving the polystyrene away with a suitable solvent. Materials containing water have been successfully replicated by using polyvinyl alcohol as the material for the negative replica. Purified crystalline virus in aqueous suspension and similar materials can be replicated by a rapid freezing technique. The frozen suspension is fractured to present a fresh surface, placed in a vacuum chamber, and warmed to −80°C until the ice matrix is etched away slightly, leaving the material to be replicated in relief. Cooling the material down again to liquid nitrogen temperatures, the surface is replicated by shadowing and a carbon substrate is evaporated on.

Limitations on observations. Specimen artifacts, instrumental defects, and image characteristics are factors in the limitations on observations. These are discussed in the following sections.

Specimen artifacts. The artifacts introduced in specimen preparation are primarily caused by the



Fig. 10. *Ultrathin section* of a portion of a mesenchymal cell from *regenerate* tissue of the newt. (*M. Salpeter*)

dehydration process *necessary* for bringing the material [illegible] sion and chemical *changes with* changing ionic concentration may be *severe*, *observations* must be supplemented with other *methods of* observation on the system being *studied* and careful evaluation made of the extent to *which* the preparation artifacts have changed the living, *dynamic* systems.

The specimen will also be altered *in the electron* microscope itself under exposure to the electron beam. Here, careful *procedures will* not, in general, cause changes in the *micromorphology* of the material. Irradiation of a *specimen* with a low intensity beam will cause chemical *changes in* the material; for example, cellulose *nitrate films* become insoluble in all solvents which *normally dissolve it.* The polymerized butyl methacrylate matrix used to embed tissue sublimes *out when thin* sections are exposed to low intensity *beams*, leaving the cytological material and enhancing *the* [illegible] of hundreds of degrees centigrade *in the* specimen material. If a thin section with the polymeric embedding *medium* is subjected directly to a high intensity *beam* the polymer will melt, causing considerable *distortion* in the section. Thermal effects can cause considerable change in some materials and care *must be* exercised in avoiding irradiation with too high intensity beams. Examination at *very high* electron optical magnifications (that is, 100,000×) *requires* very intense beams and, to minimize the *heating* effects, special illuminating systems such as those employing a double condenser are utilized.

Contamination of the specimen also occurs under exposure to the electron beam. A layer of low density, inert, carbonaceous material is laid down, sometimes at rates as high as several angstroms per second. The rate of deposition is a function of beam intensity, specimen temperature, and the amount of hydrocarbons present from pump oils and gasket grease. By surrounding the specimen with a chamber kept at −80°C the deposition of contamination can be reduced to a minimum. Special chambers achieving this effect have been designed for investigations requiring high electron current densities where the high resolution and contrast desired would soon be obliterated by the deposition of contamination.

Instrumental defects. The high-resolution electron microscopes are designed so an instrument can be adjusted to have an instrumental resolving power of better than 10 A. In an instrument so adjusted, the resolution obtained is usually limited by the resolution inherent in the specimen, especially for biological materials. But the achievement of such a high level of instrumental performance requires a high degree of skill in adjustment of the instrument by the electron microscopist. The effect which will first limit the instrumental resolving power even in a well-adjusted instrument is the deviation from axial symmetry in the objective lens, producing an astigmatic image. Correction of this asymmetry is achieved by introducing a compensating cylindrical field at right angles to the existing one, either by shims of ferromagnetic material oriented at the correct azimuth in the gap of the polepiece or by the use of an electrostatic field of correct orientation and magnitude. Externally adjustable stigmators, both of the magnetic and electrostatic type, are available which enable the operator to compensate his electron microscope and obtain resolutions down to 10 A or better.

Contrast in the electron image can be enhanced by using a stop in the objective lens limiting the angular aperture. These apertures are necessary when relatively thick specimens, such as sections and replicas, are being examined. But the objective aperture can cause a deterioration of the resolving power of the instrument, especially if asymmetric contamination forms on the aperture and causes increased astigmatism in the image. Good electron microscopy practice requires careful checking on instrumental performance using test specimens of a noncrystalline material of small particles such as carbon smoke deposited on an open specimen grid.

Image characteristics. One of the most prominent characteristics of the electron microscope is the contour effect. An instrument with a good, small effective source of illumination will produce an image in which Fresnel diffraction will be clearly observable when the objective lens is out of true focus. In practice only the first order fringe will be strongly in evidence, outlining any sharp boundaries in the image; hence the so-called contour ef

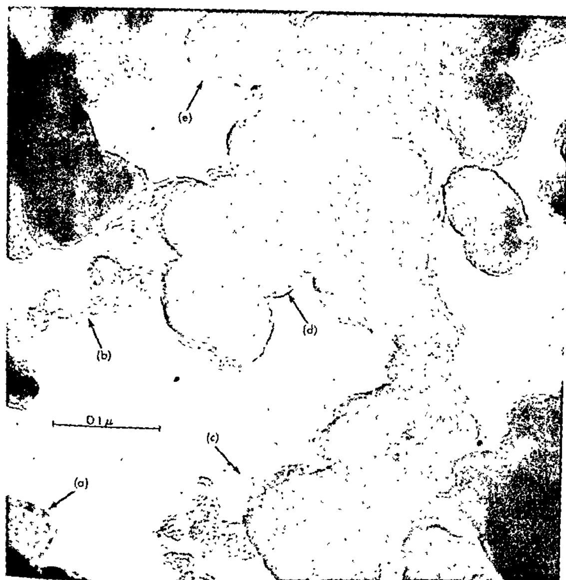


Fig. 11. A test micrograph on carbon smoke. (a) Far underfocus. (b) Slightly underfocus. (c) In focus, no Fresnel diffraction fringe. (d) Slightly overfocus. (e) Far overfocus

fect. When the objective lens current is too low so that the focal length of the lens is too long, the image is said to be underfocused. At an edge where there is a sharp change in contrast, the outside of the edge is bordered by a dark contour. In the condition of too short a focal length the image is said to be overfocused and this contour now appears inside the edge. At true focus the Fresnel fringe vanishes and a soft image is obtained. Most electron micrographs are taken at a slightly underfocus setting of the objective lens, but for critical high-resolution work a focal series should be taken. Carbon smoke is used as a test specimen. The smoke deposits in chains which project over a large depth and the different conditions of focus can be observed in a single micrograph (Fig. 11).

Crystalline material observed in the electron microscope can produce anomalous contrast effects. If a crystal is oriented so that the incident electron beam is at an angle satisfying the Bragg condition for diffraction, the crystal can scatter the total incident beam into the Bragg angle while the same crystal when not oriented at the Bragg angle will scatter a fraction of the incident beam, depending on its thickness and density. The images obtained under the two different conditions of orientation can be of very different photographic density, and care must be taken in interpreting contrast effects in observations on crystalline materials. In Fig. 12 the strong contrast effects which can be produced by the electron interference interactions when scattered from a crystal lattice are shown. Crystals



Fig. 12. Transmission electron micrograph of evaporated copper deposited on mica and transferred to a carbon substrate

of approximately uniform thickness and the observed contrast effects are caused by lattice imperfections. [B M.S.]

Bibliography: C. E. Hall, *Introduction to Electron Microscopy*, 1953; V. K. Zworykin, et al., *Electron Optics and the Electron Microscope*, 1945.

Microscope, fluorescence

A variation of the compound laboratory light microscope which is arranged to transmit ultraviolet, violet, and blue (or only the first two) radiations to a specimen. The specimen then fluoresces, that is, appears to be self-luminous and often colored. The phenomenon of fluorescence is thought to involve an electronic rearrangement in the irradiated substance. The fluorescent-antibody technique is discussed in a later section of this article. See FLUORESCENCE.

Instrumentation. Materials with absorption spectra maxima below 320 mμ require a quartz condenser in place of the usual glass condenser of the microscope. The microscope should have an aluminized front-surface mirror because silver is a less efficient reflector of ultraviolet.

The fluorescent image looks bright and has good contrast, although the amount of light is small. With weak fluorescence a monocular microscope must be used and the work is done in a darkened room. With more intense fluorescence a binocular microscope can be used. The air-glass surfaces of the microscope should be coated to decrease loss of light. Although all nonfluorescing objectives can

be used, the 8-mm and 20× ocular is often the most useful combination.

An Abbe condenser of 1.40 numerical aperture concentrates more radiation on the specimen than condensers of smaller aperture and is used for bright-field cross-filter method. Achromatic-corrected condensers may produce glare from autofluorescence of the cemented surfaces and are avoided, as are objectives with fluorite elements. Other investigators prefer the dark-field method and use a bispheric, or paraboloid dark-field condenser in place of the bright-field condenser.

The absorbing filter is usually in the microscope between the objective and the observer's eye to remove any other exciting radiation not absorbed by the specimen. This may be a nearly colorless filter absorbing only ultraviolet when the specimen is to be observed in full color, or it may be of a color complementary to the radiation used. For example, with materials absorbing blue and ultraviolet radiation a yellow filter absorbs the blue radiation beyond the specimen and passes the yellow fluorescence of the specimen to the eye. The cross-filter combination is chosen to give the best visibility for the specimen. [O.W.R.]

Fluorescent-antibody method. When infectious agents, such as viruses and bacteria, and other antigenic materials which are inanimate and principally of protein nature gain entrance into the body tissues, soluble substances are produced which specifically react with these alien materials. The soluble substances are called antibodies, and the materials which elicit their production are called antigens. When antibodies in solution are brought into contact with soluble antigen, under appropriate conditions, a precipitate is formed; or if the antigen is particulate, the particles agglutinate (stick together). In this interaction of molecules, extraordinary specificity occurs and is attributed to complementariness and close apposition of molecular shapes. See AGGLUTINATION REACTION; ANTIBODY; ANTIGEN.

Antibodies can be coupled to fluorescent dyes by gentle chemical means which do not destroy the specific reactivity of antibody with antigen. Fluorescein is the most frequently used fluorescent dye. Antibodies labeled by fluorescent dyes are called fluorescent antibodies, and these are used as immunospecific stains for the detection of antigens in cells and tissues (Fig. 1). When layered over a tissue section or cell preparation, fluorescent antibody is deposited from solution at sites of specific combination with antigen, and these regions are seen in characteristic color when the tissue section is examined with a fluorescence microscope (Fig 2). Fluorescein-labeled antibody imparts apple-green color to antigen-antibody complex, a color readily distinguished from, and rarely if ever shown by, the intrinsic fluorescence of tissue sections. In the preparation of tissue sections for study by this method, the specific activity of the antigen must be preserved, and the microscopic structure

Infectious agents studied by the fluorescent-antibody method

Viruses	
Influenza	Egypt 101
Mumps	Primary atypical pneumonia
Herpes simplex	Newcastle disease
Rabies	Fowl plague
Polio viruses	Canine distemper
Adenoviruses	Infectious canine hepatitis
Measles	Shope papilloma
Varicella (chickenpox)	Rous sarcoma
Herpes zoster	Polyoma
Psittacosis	
Vaccinia	
Yellow fever	
Bacteria	
Escherichia coli	Streptococci
Salmonella typhosa	*Pasteurella*
Shigella	*Brucella*
Neisseria gonorrhoeae	Friedlander's bacillus
Pneumococci	*Malleomyces*
Spirochetes	
Treponema	*Leptospira icterohaemorrhagiae*
Rickettsiae	
Epidemic typhus	*Coxiella burnetii*
Endemic typhus	Rocky Mountain spotted fever
Fungi	
Blastomyces dermatitidis	*Cryptococcus neoformans*
Histoplasma capsulatum	*Candida albicans*
Protozoa	
Toxoplasma gondii	*Entamoeba histolytica*
Entamoeba coli	

of the tissue should not be altered. For the most part, this requires the use of unfixed tissues which are thinly sectioned in the frozen state.

The scope of the work already undertaken with, or envisioned for, this technique establishes it as one of the major developments in microscopy and medical research. The method has been used for the microscopic identification (see table) of viruses, bacteria, rickettsiae, fungi, and protozoa in infected cells and tissues (see ANIMAL VIRUS;

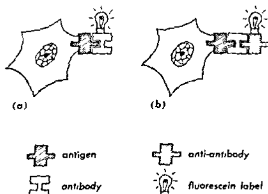


Fig. 1. Use of the fluorescent-antibody method for demonstrating (a) antigen localization in a cell, (b) antibody localization in a cell.

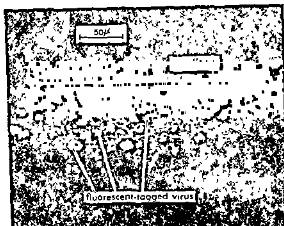


Fig. 2. Shope papilloma virus demonstrated in nuclei of epithelial cells in a virus-induced papilloma of the cottontail rabbit.

BACTERIOLOGY, MEDICAL; MYCOLOGY, MEDICAL; PARASITOLOGY, MEDICAL; RICKETTSIOSES). It has also been used for the tracing in vivo of bacterial and other foreign proteins and polysaccharides after injection into experimental animals, a study of importance for understanding the sequelae of microbial infection and disease states called allergy (hypersensitivity), and for the investigation of the cellular formation of antibodies (which are identified by staining with fluorescent antigens), a matter of fundamental concern in the problems of immunity and resistance to infectious disease. *See* BACTERIA; BIOPHYSICS; FUNGI; HYPERSENSITIVITY; IMMUNITY; POLYSACCHARIDE; PROTEIN; PROTOZOA; RICKETTSIALES. [R.C.ME.]

Bibliography: A. H. Coons and M. H. Kaplan, Localization of antigen in tissue cells: II, Improvements ir [illegible]
by means of [illegible]
91(1):1–13, [illegible]
Cytology, 2d ed., 1959.

Microscope, interference

A microscope used for visualizing and measuring differences in phase or optical path in transparent or reflecting specimens. It is closely allied to the phase contrast microscope. *See* MICROSCOPE, PHASE CONTRAST.

Simple theory. In the phase contrast microscope, the final image is produced by separating the incident wave and the wave diffracted by the object, introducing a phase difference between them, and then recombining the altered waves. The final resultant wave has a lower amplitude [illegible]

other wave which originates from the same source. In Fig. 1, A represents the incident wave, B the wave transmitted by a perfectly transparent object. Since no energy is absorbed, B has the same height as A, but is slightly delayed or altered in phase relative to A. The interfering wave is represented by the dotted line. According to the Abbe theory the final image is formed by the interference or summation of all waves that pass through the optical system. If the interfering wave and the

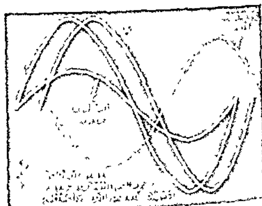


Fig. 1 Basic principle of interference microscopy A represents the incident wave, B the wave transmitted by a perfectly transparent object which has the same amplitude as A, but is delayed in phase. Interference is produced between B and another wave, shown by the dotted line, derived from the same source. The resultant differs in amplitude from wave A so that the transparent object becomes visible.

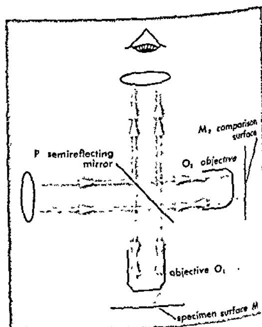


Fig. 2. The Linnik interference microscope for reflecting specimens. The light from the source is divided by the semireflecting mirror P into two beams, one of which is focused through the objective O_1 onto the specimen surface M_1, the other through the objective O_2 onto the comparison surface M_2. Interference occurs between the reflected beams which are reunited by P.

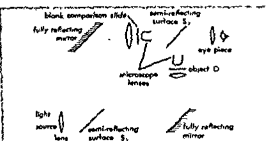


Fig. 3. A theoretical interference system for transparent specimens. The semireflecting mirror S_1 divides the light into two beams which follow similar paths. One beam passes through a microscope containing the object O, the other through a microscope containing a blank comparison slide. The beams are reunited at S_2 and can interfere.

transmitted wave B are added together algebraically, the resultant wave in Fig. 1 will have a lower amplitude than the incident wave A, so that the transparent object will now appear to absorb light. The resultant wave can be changed by altering the phase and amplitude of the interfering wave, thus enabling the appearance of the image to be varied. Since two waves can only interfere if they have the same wavelength and are derived from a common source, it is usual to divide the light from the source into two parts by means of a beam-splitting device, such as a semireflecting mirror. One beam is made to traverse the object, the other travels along a similar path, but does not pass through the

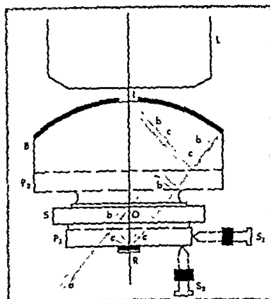
differences in optical path introduced by various parts of the object can be seen as variations in intensity or color. *See* INTERFERENCE OF WAVES.

Practical realization. The basic principles may be illustrated by reference to the Linnik microscope, which has been used for studying the surface structure of reflecting specimens (Fig. 2). Light from the source is divided at the semireflecting plate P; one beam is reflected downward through the objective O_1 and is focused on the specimen surface M_1. The transmitted beam is focused via the objective O_2 onto a flat comparison mirror M_2. Both beams are reflected back to P, where they are recombined and viewed through an eyepiece. If M_2 is tilted slightly relating to M_1, interference

finite width so that irregularities in M_1 will appear as intensity variations against a uniformly illuminated background, as in phase contrast.

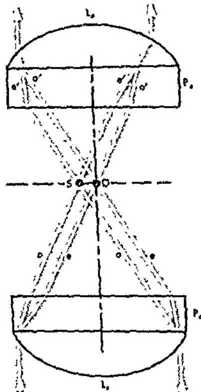
Designs for use with transmitted light are more complex. A theoretically ideal system is shown in Fig. 3, in which light from the source I is split at the semireflecting surface S_1 and sent round two different paths, one containing a microscope with the object O, the other containing a similar microscope but with a *blank* comparison slide C. The beams are recombined at the semireflecting surface S_2 and are viewed through the eyepiece E. By introducing suitable filters and transparent plates in the comparison path, the interfering beam can be varied in phase and amplitude without affecting the beam through the object. Since such a system would be expensive and difficult to operate, designs have been developed in which only a single microscope is used. A description of two such instruments in commercial production follows.

The Dyson microscope, made by Cooke, Troughton and Simms Ltd. of York, England, is shown in Fig. 4. A ray a from the substage condenser (not shown) is split at the semireflecting upper surface of a glass plate P_1 into two beams b and c. Beam b passes through the object O on the slide S, which is oil immersed to another glass plate P_2 with semireflecting surfaces. Plate P_2 is cemented to a glass hemispherical block B, the upper surface of which is metalized, except near the optical axis I. Beam c is reflected downward to a reflecting spot R, then up through the system to B and also to I. Thus, at I, interference can take place between beam b, which has passed through the object, and beam c, which has passed lateral to it. The image at I is magnified by means of an ordinary microscope objective L, to which block B is attached by a special collar. The relative amplitude of the two beams is fixed, but



Key
O = object
S = slide
B = glass hemispherical block
I = optical axis
L = microscope objective
S_1, S_2 = calibrated screws
b, c = split light beams from substage condenser
P_1, P_2 = semireflecting glass plates

Fig. 4. The Dyson interference microscope.



Key

P_e, P_o = birefringent calcite plates
L_e = condenser front lenses
L_o = objective front lenses
O = object
o = ordinary ray
e = extraordinary ray
S = point at which ordinary rays pass object
o' = ordinary ray in birefringent
e' = extraordinary ray in birefringent calcite plate

Fig. 5. The Smith-Baker microscope.

the phase difference between them can be varied by making P_1 and P_2 very slightly wedge-shaped and moving P_2 by means of a calibrated screw S_1. Interference fringes can be made to appear by rotating P_1 and P_2 relative to each other; when their wedge axes are parallel, the fringes are of infinite width, giving a uniform background as in phase contrast.

The Smith-Baker microscope is manufactured by ...

... cite plates P_e, P_o are cemented to the front lenses of the condenser L_e and objective L_o, and the microscope is illuminated with plane polarized light from a substage polarizer. Each ray is split at P_e into an ordinary ray o and an extraordinary ray e. The extraordinary rays are focused on the object O; the ordinary rays pass to one side of it at S. The two sets of rays are recombined by the birefringent plate P_o and pass through another sheet of Polaroid (not shown), which acts as an analyzer and brings the rays into the correct state of polarization for interference to occur. Thus, once again the possibility occurs of having interference between waves which have passed through the object and others which have not. The relative phase and amplitude of the interfering waves can be varied easily by rotating the polarizer and analyzer and by introducing calibrated birefringent plates known as compensators.

Measurement of phase changes. The basic principle of one method of measurement may be seen from Fig. 1. If a single uniformly refractile object occupies part of the field, the background light will be produced by wave A and the light through the object by wave B. Now, if the interfering wave is made equal in amplitude to A, but exactly out of phase with it, destructive interfer- ... that there will be no background ... of the light ... be seen by dark ground illumination. Now suppose the interfering wave is changed in phase by means of the appropriate control until it interferes destructively with B. The object will now appear dark against a lighter background. It is clear from the diagram that the phase change that has to be introduced to achieve this will be equal to that produced by the transparent object itself, that is to the phase difference between waves A and B.

Comparison with phase contrast microscopes. The phase contrast microscope can be regarded as an imperfect interference microscope in which there is incomplete separation between the two interfering beams. This results in the halo effect and the accentuation of edges. These effects can be reduced or eliminated in some interference microscopes so that accurate measurements of phase change can be made. This is not necessarily advantageous for purely observational work, however, as the optical deficiencies of phase contrast actually increase the contrast of internal details, particularly in living cells. The relative simplicity, cheapness, and ease of manipulation of the phase contrast microscope make it the instrument of choice for routine observational work and the measurement of refractive indices, but the interference microscope is essential when it is necessary to measure a phase change.

Biological applications. Apart from the considerations already mentioned, the use of the interference microscope for observing transparent specimens such as living cells calls for little comment.

The main use of the interference microscope in biology is for the determination of dry mass. There is a linear relationship between the refractive index n of a solution and the concentration C of dissolved substances, expressed by $n = n_o + \alpha C$. The term α is a characteristic constant known as the specific refraction increment whose mean value for most cellular constituents may be taken as 0.0018. If a cell is regarded as a protoplasmic jelly of refractive index n surrounded by a medium of re-



Fig. 6 (a) Fragment of glass mounted in oil, photographed with conventional illumination with reduced substage aperture. (b) Same object photographed in the Baker interference microscope under dark-ground conditions (c) Same specimen photographed with reversed contrast The interference fringes present a contour map of variations in optical path through the object. Magnification X800.

fractive index n_o, the optical path difference ϕ introduced by a region of the cell of thickness t is defined by $\phi = (n - n_o)t$; hence by combining these relationships, $\phi = \alpha Ct$. Multiplying throughout by A, the projected area of the cell, $\phi A = \alpha A ct$ is obtained. However, $Act/100$ is simply the dry mass M of the cell, so that $\phi A = 100\alpha M$ or finally $M/A = \phi/100\alpha$. In other words, the phase change ϕ is proportional to the mass per unit area of the cell. The value of the interference microscope is at present severely limited by the fact that in order to determine the total mass of a heterogeneous cell. it is necessary to make numerous measurements at every point of the cell and to integrate the result. A number of automatic integrating devices have been developed for this purpose. Since the phase change ϕ depends on the product of refractive index and thickness, one of these quantities can be calculated if the other is measured. Thus, the combination of interference microscopy and microrefractometry is capable of providing much useful quantitative information about the cell. In some cases, it is possible to determine the concentration of solids and of water, the total dry and wet mass, the cell volume and thickness. Interference microscopy can also be used for studying changes in dry mass produced by enzymic digestion or specific extraction procedures.

Nonbiological applications. Like the phase contrast microscope, the interference microscope can be applied to the study of any transparent specimen, such as fibers, crystals, and so on. A most important application is the study of reflecting specimens, such as metallic surfaces or metalized replicas of surfaces These are of special importance in metallurgy and engineering. For this type of work, it is usual to operate the instrument with fringes in the field of view; the shape of the fringe gives a contour map of the surface (Fig. 6). The sensitivity is such that irregularities of the order of 0.01 of a wavelength or even less can be observed. The method thus gives valuable information about the quality of ground, polished, or etched surfaces. *See* DIFFRACTION; RESOLVING POWER (OPTICS). [R.B.]

Bibliography: J. F. Danielli (ed.), *General Cytochemical Methods*, vol. 1, 1958; A. J. Hale, *The Interference Microscope in Biological Research*, 1958; R. C. Mellors (ed), *Analytical Cytology*, 1955, A. W. Pollister (eds.), *Physical Techniques in Biological Research*, vol. 3, 1956.

Microscope, optical

An instrument used to obtain an enlarged image of a small object. In general, a compound microscope consists of a light source, a condenser, an objective, and an ocular or eyepiece, which can be replaced by a recording device such as a photoelectric tube or a photographic plate. The optical microscope is limited by the nature of light and by the materials available for making the lenses. Some of the optical errors, for instance curvature of field and lateral color, are in general corrected by neutralizing the errors of the objective in the ocular. This article includes general discussions of microscope lenses and condensers, and specific dis sions of the following types of microsco

ordinary bright-field microscope, inverted microscope, comparison microscope, dissecting microscope, and metallurgical microscope, and a discussion of microscopy. *See* ABERRATION, OPTICAL.

LENSES

Magnifying power. The magnifying power of a compound microscope is the product of the magnification of the objective and the magnifying power of the eyepiece. The latter is computed as for a magnifier (*see* MAGNIFICATION). The magnification of the objective is equal to the distance from the second focal point to the image formed by the objective, divided by the focal length. An objective of 16-mm focal length thus has a power of 10×. It is now customary to specify objectives in terms of magnifying power instead of focal length. The distance mentioned is called the optical tube length; it is to be distinguished from the mechanical tube length, which is the length of the mechanical tube itself.

Objectives. A microscope objective consists of a set of achromatic lenses which partially or as a whole are corrected for longitudinal color, aperture errors, and asymmetry errors. The numerical aperture (NA) of the system is given by $n \sin u$, where n is the refractive index of the object space and u is the angle made by the ray of largest aperture and the axis.

A microscope objective generally consists of a collection of positive lenses comparatively close together. Color magnification and curvature of field in the objective are frequently balanced out in part by the ocular. An objective of 16-mm focal length is shown in Fig. 1*a*

For high magnifications, the first lens is plano-convex, with the convex surface either concentric or aplanatic to the aperture rays. An objective of [illegible]

[illegible] lens. Such [illegible] immersio[illegible]

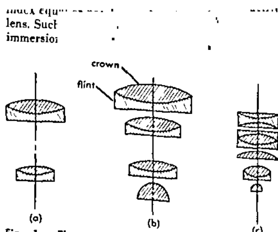


Fig. 1. Three types of microscope objective. (a) 16-mm dry achromat. (b) Typical high-power dry achromat. (c) 2-mm, NA 1.40, oil-immersion apochromat. (*Photographic Service Department, Kodak Research Lab*)

thickness separates the object from the immersion liquid for biological investigations. Figure 1 shows an immersion objective of 2-mm focal length and 1.40 NA.

Color correction. Achromatic color correction is not good enough for lenses of high aperture. The use of fluorite led to objectives corrected for more than two colors, and most achromatic lenses of high aperture are now made with fluorite. An objective whose chromatic errors are corrected for three colors and whose aperture and asymmetry errors are also corrected was called apochromatic by E. Abbe. The lateral chromatism of such an objective is great and must be removed by a special compensating eyepiece. *See* EYEPIECE.

Experiments have been made to design high-aperture microscopes for a field larger than the usual small field of 3°. This has been achieved by adding either a negative lens or a very thick positive lens having a negative Petzval sum. Such objectives are called plane-achromats. Attempts have also been made to correct field curvature in the eyepiece.

Because resolving power depends upon wavelength, objectives corrected for ultraviolet radiation have been designed. Such objectives are generally corrected for but a single wavelength and are called monochromats. They are made of quartz.

Catadioptric systems. Catadioptric systems have been developed for microscopes (*see* SCHMIDT CAMERA). A type consisting of a single element is shown in Fig. 2*a*. Their great advantage is their comparatively small chromatic aberration. Pure mirror systems would have no color aberrations. In catadioptric systems, therefore, it is customary to assign all the power to the mirror or mirrors, thus having the refracting system nearly afocal. Then the chromatic errors of the system remain small and the refracting part can be used to correct the remaining monochromatic errors. There remains, however, the disadvantage that in catadioptric systems part of the aperture is obscured by the mirror and the ensuing diffraction may damage the fine detail in the image. At present, all microscopic work in the ultraviolet region is done with catadioptric systems. A type designed for this purpose is sketched in Fig. 2*b*.

Image formation. Geometrical optics is not sufficient to explain all the details of image formation if the magnification is very high. According to geometrical optics, a point is imaged by a perfect objective as a point. But there is diffraction at the aperture, and there may also be diffraction at the object. Diffraction theory applied to the aperture shows that, because of the finite aperture of the optical system, a spherical concentric wave [illegible] produces a light pattern [illegible] distributed [illegible] faint rings [illegible] of the microscope is so high that the rings are visible, the image will contain details which are not in the object. Thus, the aperture determines the useful mag-

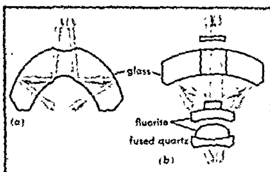


Fig. 2. Two types of catadioptric objective. (a) Maksutov type. (b) 53X, NA 0.72, ultraviolet objective, designed by Gray. Glass elements in the latter serve purely as reflectors. (Photographic Service Department, Kodak Research Lab)

nification of the microscope, and it is important that the aperture be enlarged when the magnification is increased.

Because the microscope is used to study very tiny objects and the magnification is usually high, diffraction at the aperture is more noticeable in microscopes than in other optical systems. Moreover, most objects viewed in microscopes are so small that there is a significant amount of diffraction at the object, as is seen in the case of dark-field illumination and phase microscopy.

This is of special importance if the object itself has a periodic structure. Then even a point light source will give rise to an imagelike structure arising from the interference of the rays coming from the light source in phase and being diffracted at the structured object. This gives rise to sets of images in different planes. This theory (Abbe's theory of image formation in the microscope) has been successfully carried further in F. Zernike's theory, which led to the construction of the phase microscope. *See* POINT SOURCE.

In the case of illumination with a large cone of light, the images at the Gaussian focus would fall together and thus be reinforced, so that there would only be one image plane. The results in this case, however, can also be derived from Abbe's theory if integration over the aperture is carried out. *See* MICROSCOPE, INTERFERENCE; MICROSCOPE, PHASE CONTRAST.

CONDENSERS

An external auxiliary lens is used to condense the light from a light source so that the object is brightly and uniformly illuminated. The usual purpose of a condenser system is to make sure that as much light as possible coming from all the points of the object goes through an optical system.

Condensers are used in macroscopic projection, in [illegible]
the [illegible]
mi[illegible]
light from a light source so that the cone from any object point fills most of the entrance pupil.

Condenser system. A condenser system is usually arranged to image the light source onto the entrance pupil of the optical system (Köhler illumination). The condenser is generally corrected for spherical aberration, color, and sine condition, although the requirements are slightly different than in an image-forming system. Condensers frequently consist of a number of planoconvex lenses with the plane side towards the objective. Sometimes one surface is made aspheric to improve the light concentration. Condensers for projection optics are rarely achromatized, but the effect of color magnification is decreased by vignetting the colored borders. For microscope substage condensers, however, achromatism is a necessary requirement.

Aperture. The aperture of the condenser must be at least as large as that of the objective with which it is used. Because microscope objectives are generally designed in such a way that they are excellently corrected for color, aperture, and asymmetry errors only for about ⅞ of their aperture, the condenser need fill only this much of the entrance pupil. A good test of whether the condenser of a microscope is well adjusted is to remove the object and ocular and see whether the exit pupil of the objective is filled uniformly with light up to ⅞ of its aperture.

Dark field. Thus far, only condensers for giving a bright field have been discussed. In microscope practice, it is sometimes advantageous to increase the contrast of a group of small objects by making them appear as bright objects on a dark background. This is achieved by arranging the condenser so that the direct light is cut off by a stop and only the light diffracted from the object enters the microscope. The cardioid condenser sketched in Fig. 3 is an example of such a dark-field condenser.

The principles outlined here have been elaborated in later years in phase microscopy (*see* MI-

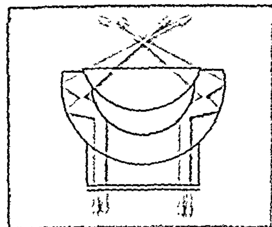


Fig. 3. Cardioid condenser. Unshaded meniscus area is an airspace; other unshaded areas represent the portions of the condenser through which light passes. (Photographic Service Department, Kodak Resear[illegible] Lab)

CROSCOPE). In the phase microscope, an annular diaphragm is placed at the front focus of the substage condenser so that the object is illuminated by a hollow cone of parallel light. A phase plate is put at the image of the diaphragm formed by the objective. This phase plate usually consists of a transparent annular layer evaporated on a transparent plate, the added layer corresponding in size and shape to the image of the diaphragm and having a thickness equal to ¼ of a selected wave-

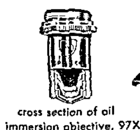
optical and mechanical features of the microscope



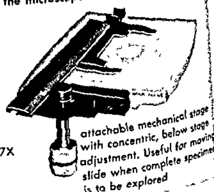
cross section of low power objective, 10X



cross section of "high dry" objective, 43X



cross section of oil immersion objective, 97X



attachable mechanical stage with concentric, below stage adjustment. Useful for moving slide when complete specimen is to be explored

Fig 4. Diagram of light microscope. (American Optical Company, Instrument Division)

length of light. A metallic layer is generally superimposed to diminish the transmittance of the annulus. When diffraction takes place at the object, the central maximum has a phase shift of $\lambda/2$ with respect to the diffracted light. This phase shift results in an intensity difference in the image which increases the contrast. [M.H.]

TYPES OF MICROSCOPE

The types of microscope discussed in this section are variations of the light- or bright-field microscope.

Light microscope. The mirror, condenser, oculars, and body tube of the light microscope are frequently known as the optical train. The stand, stage, and adjustments comprise the mechanical part of the microscope (Fig. 4).

A mirror is usually attached to the substage of the microscope to reflect light along the optic axis of the microscope. When no condenser is used, the concave mirror is used because it concentrates more light on the specimen; a plane mirror is used with a condenser.

The condenser concentrates light onto the specimen at an angle to fill the objective. Laboratory microscopes are usually supplied with an uncorrected 2-lens Abbe condenser with NA 1.25, consisting of a double convex lower and a hyperhemisphere upper lens. A 1.40 NA Abbe is useful for concentrating radiation for fluorescence microscopy. Well-corrected objectives require aplanatic condensers corrected for chromatic aberration for efficient observation.

The objective is the basic part of the microscope; it forms the image that is again enlarged by the eyepiece. Objectives vary from a simple doublet lens to complex corrected lens systems. Achromatic objectives are corrected for spherical aberration in one color and for chromatic aberrations in two colors. Apochromatic objectives are corrected to focus three colors together and the spherical aberration is minimized for two colors. Some semiapochromatic and fluorite objectives are of intermediate correction. Fluorite objectives include a lens of crystal fluorite. *See* CHROMATIC ABERRATION

As previously indicated, one important measure of an objective is the numerical aperture. With air between the specimen and the objective the maximum NA is about 0.92. Water-immersion objectives have a greater NA, and with immersion oil, an NA of 1.4 is available. Objectives with an NA of 1.6 have been made but require special immersion and mounting media. The resolving power of an objective, the least distance at which two objects can be seen to be separate, is equal to the wavelength of light λ divided by the sum of the numerical apertures of the condenser and objective used. The larger the numerical aperture, the greater is the resolving power. The depth of field seen in focus at one time and the working distance of objective decrease with an increase in the numerical aperture. The light passed through a microscope is proportional to the square of the numerical aperture and to the inverse of the square of the magnification. The numerical aperture is engraved on the objective and can be measured with an appotometer. Objectives are described also by the equivalent focal length. Objectives of shorter focal length have less depth of field, less working distance, and greater magnification.

Photomicrographic objectives are designed to produce a flat image with little distortion. Some objectives obtain a flatter field by means of a concave rather than a flat front lens. Apochromatic objectives are under-corrected and compensating oculars must be used with them to complete the correction for color and for best resolution.

For convenience, 2–5 objectives can be mounted on a revolving nosepiece to be parfocal and parcentric, so that the specimen remains almost in focus at the center of the field as the objectives are changed. For more critical work, utilizing interference and polarizing microscopes, individually

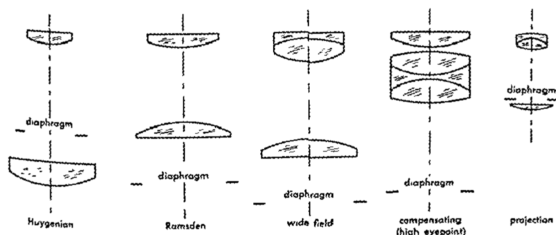


Fig 5 Types of ocular.

adjustable quick-change nosepieces are employed.

The commonly used Huygenian ocular has a fairly flat field with marked pincushion distortion (Fig. 5). Compensating oculars complete the color correction for apochromatic objectives and have less distortion, but they do have curvature of field. To obtain a flat field with minimum distortion for microprojection or photomicrography special projection oculars are designed, one type with a color-corrected minus lens called a negative amplifier, and the other a positive-projection ocular with a focusable eye lens. Oculars with a high exit pupil are useful for spectacle wearers. Other oculars are designed to give a wide field of view.

The monocular body tube may be of adjustable length. American microscopes are designed for a mechanical tube length of 160

to correct for the spherical aberrations from coverglasses of incorrect thickness.

Binocular bodies are designed for the use of both eyes. Most binocular bodies use prisms to reflect one-half of the light to each eye. Because each eye sees the same field, these binocular bodies do not give stereoscopic vision. The binocular is often longer than the monocular body and the proper tube length is maintained with a compensating lens. Because the tube length is fixed, it is essential for critical microscopy with binocular bodies to use coverglasses 0.18 mm thick. Binocular bodies can be made with stops, or polarizing materials, so that each eye sees with the corresponding half-aperture of the objective and stereoscopic vision is possible, *but at one-half* the resolving power that would have been obtained with full aperture.

Trinocular bodies are binocular bodies with a third tube for a camera.

The lowest magnification which will resolve the detail required should be used. High magnification gives images that are less bright and therefore difficult to see. Magnification greater than that required for complete visible detail is called empty magnification and is useless for seeing, although sometimes helpful for measurement. For visual microscopy the total magnification (magnification of the objective times that of the ocular) should be about 1000 times the numerical aperture in use. Apochromatic objectives and compensating oculars are desirable for best vision and photomicrography in color. The research microscope has a larger, heavier stand with more precise adjustments for convenience and measurements. For special applications, microscopes with phase, interference, polarizing, and other special equipment are available.

Inverted microscope. The inverted microscope has the body of the microscope, including the objective and the ocular, below the stage and the illumination above the stage for transmitted light. With opaque materials, the vertical illuminator is used under the stage near the objective. The inverted microscope is especially useful for the ex-

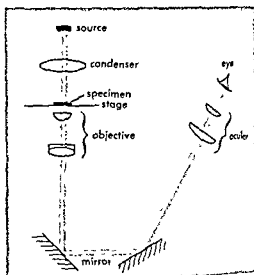


Fig. 6. Optical diagram of inverted microscope

amination of surfaces (Fig. 6). Specimens placed on the stage can be held substantially in focus. Large and awkward specimens can be moved over the stage more readily than with the usual microscope. The inverted microscope is also useful for microdissection and the observation of hanging-drop preparations, and is convenient for observing chemical reactions, melting-point determinations, and photomicrography. The camera can be included in the base, as in the metallograph microscope, for stability. Either monocular, or binocular bodies can be used with the inverted microscope.

Comparison microscope. The comparison microscope is an arrangement of two microscopes connected by a special viewing ocular so that the field of one microscope is seen at one side of a vertical dividing line and the field of the other microscope on the opposite side of the dividing line; or it may be a projection type of microscope in which the image is compared with a template or known pattern.

When two microscopes and a comparison ocular are used, the magnifications of each must be matched. The specimens are placed on the microscopes and usually require separate lighting. A common application is the examination of bullets. A test bullet fired from the suspected gun is placed under one microscope and the bullet recovered from the scene of the crime is placed under the other microscope. Holders which allow rotation of the bullets are turned and if the grooves on the two bullets match, it is evident that they came from the same gun. Comparison microscopes can be used to compare grain size, structure, distribution of elements, color, and various other characteristics of any two specimens. Either transmitted light, vertical illumination, or a combination of lighting types can be used.

A projection microscope with a built-in screen can also be used as a comparison microscope. The image of the part is focused on the screen and the contours of the image are measured with scales, or

compared with a template of the desired form, for example, a profile of a gear, or a screw thread. The stage of the microscope is usually modified to have a proper holder for the kind of specimen to be examined.

Dissecting microscope. Dissecting microscopes are of two types. The simplest is a magnifying glass mounted on a support above a glass plate, used for the dissection of materials. A mirror may be present to reflect light to the specimen

The more usual dissecting microscope, often called a Greenough microscope, is a stereoscopic microscope composed of two separate microscopes fastened and used together as a single unit on one stand (Fig. 7).

This is a truly stereoscopic instrument because the right eye sees the specimen from the right side and the left eye from the left side. Prisms are usually included in the body tube to erect the image; thus movements of the specimen are direct and are not reversed as with the monobjective microscope. Because two objectives are required, the mechanical difficulty of placing the lenses close together limits the numerical aperture to about 0.12. There is no advantage in using more than 120 diameters magnification, because further magnification would be empty or useless.

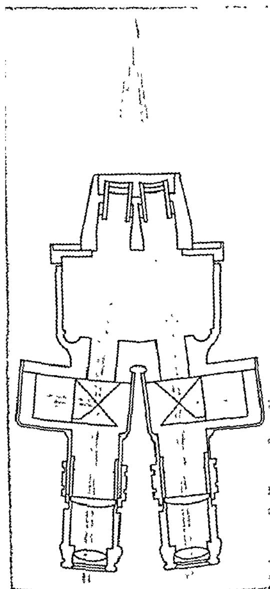


Fig. 7. Light path of binocular biobjective mic

The binocular microscope was developed for biological dissecting, but with the need for small parts in industry it is now used extensively for the examination and assembly of small parts such as transistors The body and focusing adjustment can be mounted on lathes and other machinery when the work must be controlled visually to close tolerances, or they can be mounted on a stand with a long arm for examination of large specimens, or to assist the surgeon in the operating room.

When the angle of the objectives and oculars is the same, the specimen is seen in true depth. By changing these angles the perception of depth can be increased or decreased. Hyper- or increased stereoscopy is useful for biological dissection.

Examination with the stereoscopic microscope is helpful in orienting the microscopist to the specimen before instruments of greater magnification are used By proper illumination both opaque and transmitted materials can be examined.

Metallurgical microscope. The metallurgical microscope is a laboratory microscope with a focusing stage and a vertical illuminator, used primarily for the examination of metal surfaces (Fig. 8).

The specimens are usually imbedded in molded plastic so that the surface is at a definite position, surfaced with a series of increasingly finer abrasives, and polished. To differentiate the constituents of the metal the surface is etched lightly by chemical treatment before examination with the microscope.

The metallurgical microscope shows the structure of the metal, including grain size, as well as the nature and distribution of the components. The roughness or polishability of the metal can be studied. Because many of the metals used commercially are mixtures, or alloys, rather than pure metals, the metallurgical microscope is important for analysis and for assessing the effects of heat treatment and surface changes.

MICROSCOPY

This discussion of microscopy considers the following with relation to the optical microscope: illumination, calibration and measurement, immersion fluids, mountants, and the hanging-drop method.

Illumination. Illumination of the microscope is obtained from a bright surface or from a luminous source concentrated with the aid of condensing lenses Sources commonly used are tungsten-filament lamps and carbon, mercury, and xenon arcs.

With an illuminated-surface light source, the lamp is positioned and the microscope condenser

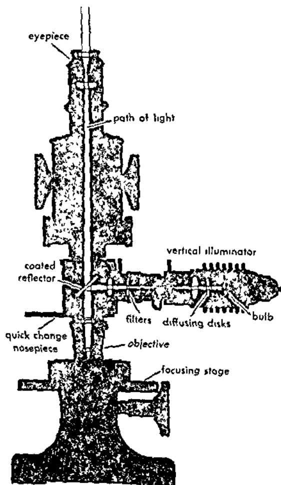


Fig. 8. Light path of a metallurgical microscope with a vertical illuminator (*American Optical Company, Instrument Division*)

focused so that the field and aperture of the microscope are lighted as uniformly as possible.

Illuminators with focusing condensers and iris diaphragms, called research lamps, are used for *critical*, Köhler, and other more efficient methods.

Critical illumination. In critical illumination an *image of* the light source is focused on the *specimen from* a uniform source, for example, a ribbon-filament lamp. This method is no more and no less critical than other methods, but the name has evolved and is so used (Fig. 9).

Parallel light from an illuminator *focused to infinity* is directed into the microscope *condenser*. The microscope condenser is then focused *so that* the image of the source is in the plane of *the specimen*.

Köhler's method. This method is used with coiled filaments or other sources of *irregular* form or brightness to obtain a uniformly illuminated *field*. An image of the filament large enough to fill the opening of the iris is focused on the microscope condenser. The *microscope* condenser then is focused so that the *image* of the iris diaphragm on the lamp is in *focus with* the specimen and the lamp iris is opened only enough to fill the field of view. The iris of the microscope condenser is opened only enough to illuminate the back aperture of the objective. No ground glass is used. The condensing lens of the illuminator becomes the effective source and the field is uniformly illuminated, even though the source itself is not uniform. Loss of contrast from misplaced glare light is minimized (Fig. 10).

Shillaber's type 3. This method is useful when the lens systems of the condensers are inadequate to illuminate the field and aperture of the microscope.

A ground glass is placed close to the condensing lens of the illuminator between the lens and the lamp bulb, or the surface of the lamp is ground on the side facing the lens. Because the ground surface does not diffuse all the light, the condensing lens can concentrate more light into the microscope than could be obtained from the same area of a surface-type illuminator. An effectively larger source results which is useful with low-power (searcher) objectives for a large field of view and when the lamp filament is not large enough for the condensing system, for example, some sources built into the base of the microscope. This method is preferable when a diffuser must be used, but more stray or glare light is produced than with the Kohler method.

Oblique illumination. Oblique illumination occurs when the mirror is moved to one side of the optical axis of the microscope, the nearly closed iris diaphragm is moved away from the axis of the

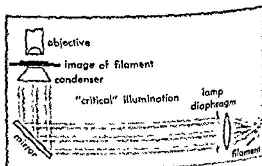


Fig. 9. Light path with critical illumination (*American Optical Company, Instrument Division*)

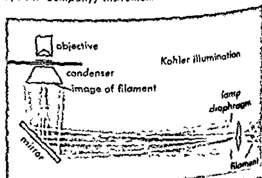


Fig. 10. Light path with Köhler illumination. (*American Optical Company, Instrument Division*)

condenser, a diaphragm is moved across the condenser aperture, or part of the exit pupil (Ramsden disk) is covered. A plastic, pseudo-three-dimensional appearance may occur, and parallel detail in the specimen is emphasized with this one-sided lighting. Interpretation of what is seen is difficult.

Vertical illumination. Vertical illumination uses a partly silvered mirror, or a prism, to reflect light through the objective onto the specimen. The light reflected back through the objective from the specimen forms the image which is seen.

Epi-illumination is vertical illumination with the illuminating light passing around a special objective to the specimen and only the light reflected from the specimen passing through the objective. Better definition results from the separation of the paths of the illuminating and viewing light.

Metals, ores, minerals, and opaque materials with adequate reflecting surfaces are examined with vertical illumination.

Dark-field illumination. In dark-field illumination the specimen appears bright, or self-luminous, against a black background. The illuminating beam is a hollow cone of light formed by an opaque stop at the center of the condenser, large enough to prevent any direct light from entering the objective. A specimen placed at the concentration of the light cone is seen with the light scattered or diffracted by it and the smallest particle revealable depends upon the intensity of the available light, even though the particle may be too small for resolution as to size and shape. Size can sometimes be inferred from the number of particles found in a given volume of the specimen (Fig. 11). Ultramicroscopy and Rheinberg illumination are modifications of dark-field illumination.

Ultramicroscopy, used for the examination of colloids and smokes, is a dark field obtained with an intense narrow beam of light directed through the specimen at right angles to the optical axis of the microscope.

Rheinberg illumination or optical staining is a modification of the dark-field method. The central disk is transparent and colored, rather than opaque, and an annulus of a complementary color

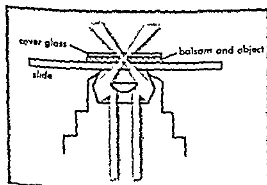


Fig. 11. Dark-field illumination. (American Optical Company, Instrument Division)

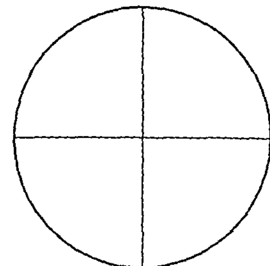


Fig. 12. Cross-hair reticule. (American Optical Company, Instrument Division)

fills the remaining condenser aperture. The specimen is seen in the color of the annulus against a background of the color of the central disk; for example, when the annulus is yellow and the background blue, the specimen appears yellow against a blue background.

Modification by filters. Filters are used between the microscope illuminator and the microscope to control the intensity or quality of the illumination. Filters can be liquids, in a flat-sided container, or solids. Solid filters are made of colored glass, gelatin, or other materials. The type of filter used depends on the requirements of the examination.

Clear filters of water or of heat-absorbing glass are used to remove the excess heat from the lamp beam when delicate specimens must be protected. Blue filters remove the excess yellow from tungsten light so that it resembles daylight. Neutral filters of glass or metal (Inconel) deposited on glass are used to reduce the amount of light without altering the color of the light.

Colored filters change the quality of the light by selectively absorbing certain wavelengths and are grouped into broad- and narrow-band types. Broad-band filters are used to increase visibility by modifying the color contrast of the specimen, for example, a yellow (minus blue) filter transmits all of the spectrum except the blue. Complementary filters increase contrast, and filters of similar color decrease color contrast. Polychromatic filters, transmitting two or more spectral colors (for example, a minus green filter passing blue and red) are useful with some stained specimens. Smaller regions of the spectrum are isolated with narrow band-pass filters, interference filters, and by combinations of filters. Monochromatic light usually is obtained from a single spectral line of an arc source rather than with filters.

Calibration and measurement. The size of the object under examination is frequently a desirable

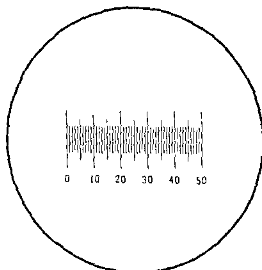


Fig. 13. *Scale reticule, 9 times normal size. (American Optical Company, Instrument Division)*

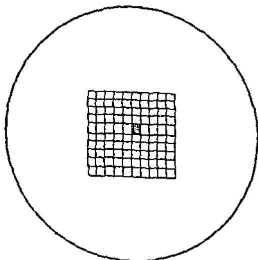


Fig 14. *Net reticule. (American Optical Company, Instrument Division)*

criterion for identification. A reference reticule is placed on the diaphragm in the focal plane of the ocular and seen in focus with the specimen from measurement in microscopy. A cross-hair reticule (Fig. 12) is satisfactory for position; scales (Fig. 13) are used for measuring distance and a net reticule (Fig. 14) for counting and drawing. Oculars with a focusing eye lens permit focusing the scale to the microscopist's eye.

Angular measurements are made by orienting one side of the specimen to the cross hair, reading the position of the graduated stage of the microscope, and rotating the stage until the other side is in line with the cross hair. The difference between the first and second readings of the scale gives the angle. Another method uses a goniometer ocular and measures the angle with the scale of the goniometer.

Linear measurements are made by placing one edge of the specimen at the cross hair, reading the position of the graduated mechanical stage, moving

the readings is the distance. This method is useful for larger specimens because the mechanical stages are ordinarily not graduated closer than to 0.1 mm. For measurements of smaller specimens a scale is used in the eyepiece and the size of the specimen obtained from the scale (Fig. 13). The actual or true distance of the ocular scale depends upon the magnification of the microscope and is obtained by calibration with the aid of a stage micrometer. The true distance x seen on the stage micrometer which corresponds to the number of divisions y of the eyepiece micrometer is noted. Dividing this true distance by the number of divisions in the eyepiece micrometer gives the distance $c = x/y$ each one subtends. The number of eyepiece divisions covered by the specimen multiplied by the calibration constant c gives the actual length of the specimen (Fig. 15). Change of eyepiece, objective, magnification, or tube length of the microscope necessitates recalibration.

More precise measurements can be made with a screw micrometer eyepiece, a filar micrometer, or a step micrometer eyepiece. The movable scales with graduated controls facilitate measurement, but must also be calibrated (Fig. 16).

Immersion fluids. Air or other low-refractive-index material between the specimen and the objective of the microscope limits the angle of light that can enter the objective, the numerical aperture, and the resolving power of the system. To obtain greater values, a medium of higher refractive index must be placed between the condenser of the microscope and the microscope slide, and between the cover slip of the microscope and the

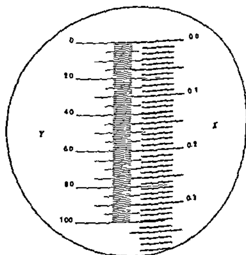


Fig. 15. *Calibration of eyepiece scale y, with a stage micrometer x. (American Optical Company, Instrument Division)*

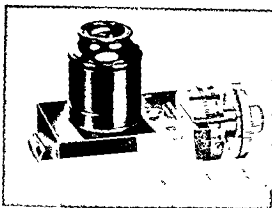


Fig. 16. Screw micrometer ocular. (American Optical Company, Instrument Division)

front lens of the objective. Immersion fluids may be aqueous (such as water or glycerin), oils, or organic materials.

In addition to the correct refractive index and dispersion, a satisfactory immersion oil must possess the following physical properties: chemical inertness, no tendency to spread or creep, no hardening when exposed to air, and little change with age.

Mountants. Mountants for specimens to be observed under the microscope are usually classed as temporary or permanent. Temporary mountants are water, physiological saline, serum, and the like. Such preparations last longer when the cover glass is sealed to the slide with petroleum jelly or paraffin. Semipermanent mounts are made with glycerin, glycerin jelly, or oils, and permanent mountants from gums, resins, or polymers which harden and remain solid at room temperatures. Canada balsam and damar are classic mountant materials but have been largely replaced by synthetic polymers, such as Clarite, and polyvinyl alcohol.

A good mountant should provide enough difference in refractive index for contrast within the specimen, be transparent, not change color or consistency with time, hold the cover glass firmly in place, and not fade stains, or otherwise alter the specimen. For the large number of mountants used in microscopy, see MICROTECHNIQUE.

Hanging-drop preparations. The specimen is placed in a drop of a suitable fluid on a cover slip and the cover slip is inverted over a slide which has a hollow ground in it (Fig. 17). Various concavities are used—round, flat, and so on—and the thickness of the slide varies with the preparation. The coverglass is usually sealed to the slide to retard evaporation. More complex preparations have channels so that the environment of the specimen can be changed or kept constant. Some hanging-drop-preparation slides are open at the bottom to make it possible to bring microdissection needles to the specimen for surgical procedures on the specimen while it is observed through the microscope. See MICROMANIPULATION. [O.W.R.]

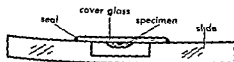


Fig 17. Hanging-drop preparation.

Bibliography: R. Barer, A. Bouwers, and D. S. Gray, Reflecting microscope, *Proc. London Conf. Opt. Instr.*, pp. 43–76, 1951; H. Boegehold, *Das Optische System des Mikroskops*, 1958; A. Bou-

pany, *Photomicrography*, 14th ed., 1944; P. Gray, *The Microtomist's Formulary and Guide*, 1954; M Herzberger, *J. Opt. Soc. Am.*, 1936; R. M. Jones (ed), *Handbook of Microscopical Technique for Workers in Animal and Plant Tissues*, 3d ed., 1950; O Lummer and F. Reiche, *Die Lehre von der Bildentstehung im Mikroskop, von E. Abbe*, 1910; C. P. Shillaber, *Photomicrography in Theory and Practice*, 1944; J. Strong, *Concepts of Classical Optics*, 1958.

Microscope, phase contrast

A microscope used for making visible differences in phase or optical path in transparent or reflecting specimens. It is one of the most important instruments available for studying living cells and is widely used in biological and medical research.

Microscopy of transparent objects. When a light wave passes through an absorbing object, it is reduced in amplitude and intensity. Since the human eye and the photographic plate are sensitive to variations in intensity, such an object will give a visible image when viewed through an ordinary microscope. A perfectly transparent object does not absorb light, so that the intensity remains unaltered, and such an object is essentially invisible in an ordinary microscope. However, the light that has passed through the transparent object is slowed down by it and arrives at the eye a minute fraction of a second later than it would otherwise have done. Such delays, or in technical language, phase changes or differences in optical path, are not detected by the eye, and the fundamental problem in the microscopy of transparent objects, of which living cells are important examples, is to convert them into visible intensity changes. For further discussions of basic principles see ABSORPTION (ELECTROMAGNETIC RADIATION); DIFFRACTION; INTERFERENCE OF WAVES; LIGHT.

Simple theory. A consideration follows of what happens when a light wave passes through a partially absorbing object, such as a stained biological specimen. In Fig. 1, let A be the incident wave. When the wave passes through the object, energy is absorbed so that the transmitted wave B is reduced in height or amplitude. The intensity is proportional to the square of the amplitude. Wave B can also be represented as the sum of the inciden

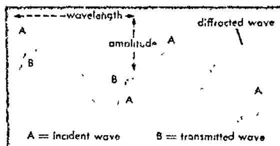


Fig. 1. Effect of a partially absorbing object on a light wave. The diffracted wave is half a wavelength out of phase with the incident wave

wave A and another wave shown by a dotted line. This wave has a real physical existence and represents the light scattered or diffracted by the object. The German physicist E. Abbe was the first to stress the importance of diffraction in image formation and showed that the final image was formed by interference or addition of the incident and diffracted light. If the incident wave and the diffracted wave are added together algebraically, that is, upward displacements being treated as positive and downward ones as negative, the resultant is wave B. In the case of a partially absorbing object, therefore, the trough of the diffracted wave coincides with the crest of the incident wave, that is,

height of the wave

wave B in Fig. 2. Once again, wave B can be represented as the sum of the incident wave A and a diffracted wave (Fig. 2*b*). This diffracted wave is no longer half a wavelength out of phase with wave A. If wave B is only slightly delayed, the diffracted wave is approximately one-quarter of a wavelength

difference between

them could be increased from about one-quarter of a wavelength to half a wavelength, as in Fig 2c. The crest of one wave would now coincide with the trough of the other, and the new resultant wave would be indistinguishable from wave B in Fig. 1. In other words, the invisible transparent object would be indistinguishable from a partially absorbing object. This was first realized by Frits Zernike, who was awarded the 1953 Nobel Prize in Physics for his invention of phase contrast.

Practical realization. The essential features of a phase contrast microscope are shown in Fig. 3. The practical problem is to find some way of separating the incident or direct light from that diffracted by the object. This is done by placing a diaphragm D of easily recognizable shape, such as an annulus, at the front focal plane of the substage condenser C. Light from each point of the focal plane passes as a parallel pencil of rays through the specimen S and is brought to a focus at the rear focal plane P of the objective O. Thus, on removing the eyepiece, an image of the annulus will be seen at the back of the objective lens. This image corresponds to the incident light. In addition, when a specimen is present, some light is diffracted by it and spreads out to fill the whole of the back lens of the objective. Thus, apart from the small area of overlap over the image of the annulus, the direct and diffracted waves are essentially separated at the plane P. A phase plate is now inserted at this level. This can be a transparent disk with an annular groove of such dimensions that it coincides exactly with the image of the diaphragm D. All the direct light now passes through the groove in the phase plate, whereas the diffracted light passes mainly outside the groove. Since the diffracted light has to pass through a greater thickness of transparent material than the direct light, a phase difference, depending on the refractive index of the phase plate material and on the thickness of the groove, is introduced between them. If this phase difference is about one-quarter of a wavelength, the basic conditions for phase contrast will have been achieved (Fig. 2).

In practice, a partially absorbing layer of metal or dye is usually deposited over the phase plate an

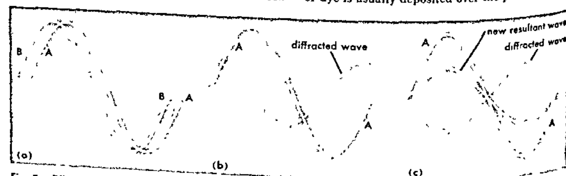


Fig 2. Effect of a transparent object on a light wave. (a) The incident wave A is unaltered in amplitude but delayed in phase and is transmitted as wave B. (b) B can again be represented as the sum of the incident wave A and a diffracted wave. The diffracted wave is no longer half a wavelength out of phase with A, but if the phase difference can be changed to half a wavelength (c) the new resultant wave will resemble that produced by a partially absorbing object.

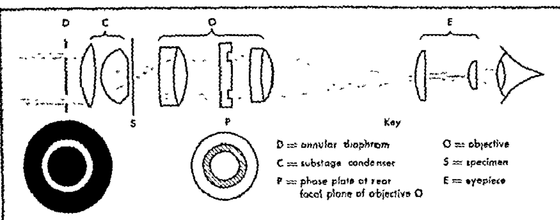


Fig. 3. Diagram of a phase contrast microscope.

nulus. Figure 2c shows that the amplitude of the direct wave is usually greater than that of the diffracted wave so that, although the resultant wave is reduced in amplitude, it is still far from zero, and a transparent object will not appear perfectly black, but gray. If the amplitude of the direct wave is reduced after it has passed through the object, it can be made equal to that of the diffracted wave so that the resultant is zero, and the object appears perfectly black. Unfortunately, in order to match the amplitude of all possible diffracted waves, it would be necessary to make a variable-absorption phase plate. Such a device is expensive, and in practice most manufacturers use fixed phase plates with absorptions of between 50 and 80% Each objective must have a corresponding substage diaphragm. The commonest arrangement is a rotatable wheel in the substage condenser containing three or four diaphragms of different sizes and also a clear circular aperture which enables the instrument to be used with conventional illumination.

Image interpretation. The phase change or optical path difference ϕ introduced by an object is defined by the relationship $\phi = (n_o - n_m)t$, where n_o is the refractive index of the object, n_m that of the surrounding medium, and t the object thickness; ϕ is usually expressed either in terms of the wavelength, in which case t must also be expressed in the same unit, or in angular measure In the latter notation, one wavelength is equal to 360° or 2π radians. Although the phase contrast microscope converts variations in ϕ into variations in intensity, the relationship between these quantities is not a simple one. In general, as ϕ increases, the image becomes darker; but beyond a certain value of ϕ, it becomes lighter again, and finally, for very large values of ϕ, the contrast actually becomes reversed, so that the image becomes brighter than the background. Thus, it cannot be assumed that because one part of the image appears darker than another, it will necessarily correspond to a region of greater optical path. Even these basic relationships are disturbed because the theoretical requirement of complete separation between the direct and diffracted light can never be achieved in practice. This results in the appearance of a bright halo around every dark detail and an accentuation of edges and sharp discontinuities in the object. Although these effects do not allow the instrument to be used for quantitative measurements of optical path, they are not really disadvantageous for purely observational work.

Biological applications. The phase contrast microscope is the routine instrument for the examination of living cells. It is now possible to study the structure of living cells under excellent optical conditions and with no loss in resolving power. Accurate observations have thus become much easier to make, and in particular, the use of phase contrast cinemicrography has made it possible to study changes in cell structure during such processes as the movement and division of cells with much greater clarity than hitherto. The method is also useful for the study of unstained tissue sections and is finding considerable use for the comparison of material in the electron and optical microscopes.

A quantitative application is microrefractometry. It follows from the basic definition of phase change, namely $\phi = (n_o - n_m)t$, that if the refractive index n_m of the mounting medium is made equal to that of the object n_o, ϕ becomes zero irrespective of the object thickness t, and since the intensity is a function of ϕ, the object will become invisible. The phase contrast microscope can thus be used as a very sensitive null indicator for measuring refractive indices. The principle is to immerse the object in a series of media of graded refractive index until one is found that makes the object invisible. This is one of the most sensitive methods of microrefractometry now available, and it can be extended to the quantitative refractometry of living cells. In this case, a suitable nontoxic medium must be used. The most suitable medium is a concentrated protein solution such as Armour's bovine plasma albumin fraction V, with added salts (see Fig. 4). The importance of cell refractometry rests on the fact that there is a linear relati ship between the refractive index n of a sol



Fig. 4. (Left) Phase contrast photograph of living locust spermatocyte in a saline medium. The refractive index of the cell is so much higher than that of the medium that all internal detail is obscured. (Right) A similar cell immersed in 9% bovine plasma albumin solution. The refractive index difference is much less and the internal nucleoplasm and chromosomes can be clearly seen.

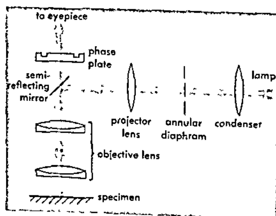


Fig. 5. Phase contrast microscope with vertical illumination for reflecting specimens.

and the concentration C of the dissolved substance. Thus, $n = n_0 + \alpha C$ where n_0 is the refractive index of the solvent for example, water and α is a characteristic constant known as the specific refraction increment. The mean value of α for most cellular constituents may be taken as .0018, and the for-

(see

study of trans-parent specimens such as crystals and fibers calls for little comment except to point out that if the specimen is birefringent, a single polarizer must be used in order to avoid a confused image.

Industrial applications include the examination of chemicals, oils, waxes, soaps, paints, foods, plastics, rubber, resins

co

im naturally reflecting, or a thin film of reflecting material must be deposited on it or on a surface replica. A typical arrangement is shown in Fig. 5. A condenser lens C forms an image of a lamp X on an annular diaphragm D. A semireflecting cover slip is placed at S, just above the objective lens O. The projector lens L and objective O combine to form an image of D at P, above S. A phase plate is inserted at P (compare Fig. 3). Normally, it is desirable to place the phase plate at the rear focal plane of the objective; but if this were done here, the incident light would have to pass through the metalized part of the phase plate and would lose intensity. Stray light and glare would also occur because of reflections at the surface of the phase plate. Apart from these modifications, the system is similar to that for phase contrast with transmitted light. Differences in height in the reflecting specimen M will produce differences in optical path length, and these will be converted into intensity differences.

[A B]

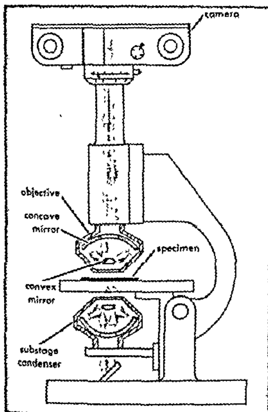
Bibliography: A. H. Bennett, H. Jupnik, H. Osterberg, O. W. Richards, *Phase Microscopy*, 1951; R. C. Mellors (ed.), *Analytical Cytology*, 1955; G. Oster and A. W. Pollister (eds.), *Physical Techniques in Biological Research*, vol. 3, 1956.

Microscope, reflecting

A microscope whose objective is composed of two mirrors, one convex and the other concave. The imaging properties are independent of the wavelength of light, and this freedom from chromatic aberration allows the objective to be used even for infrared and ultraviolet radiation. Although the reflecting microscope is simple in appearance, the construction tolerances are so small and so difficult to achieve that the system is used only when refracting objectives are unsuitable. The distance from the objective to the specimen can be made very large; this large working distance is useful in special applications, such as examining objects situated within metallurgical furnaces. Reflecting microscopes have been mainly used for microspectrometry in the infrared and the ultraviolet, and for ultraviolet microphotography.

Optical system. The concave mirror collects light from the specimen under examination. The convex mirror intercepts this light just above the point where the rays would be focused by the concave mirror, and redirects the light up the microscope tube to the plane where the primary image is formed (see illustration). The curvatures and spacing of the mirrors can be so chosen that the optical aberrations of the mirrors are mutually compensating, provided that the numerical aperture does not exceed about 0.60 (visible light). The limiting aperture is somewhat smaller than this for ultraviolet radiation and somewhat larger for infrared, and is smaller for large working distances than for small working distances.

The optical designer, in attempting to achieve numerical apertures larger than about 0.60, may resort to mirrors that depart slightly from spherical curvature; alternatively, he may add one or more lens elements. By such means, objectives of numer-



Reflecting microscope arranged for photomicrography

ical aperture greater than unity and free of chromatic aberration can be realized.

The convex mirror unavoidably obscures a small central portion of the aperture. This alters the diffraction of light (upon which the ultimate resolving power of the objective depends) but fortunately this effect is negligible if proper care is used in the design and construction.

Substage illumination of the specimen can sometimes be achieved by a conventional refracting condenser if only visible light is used. Usually, however, the chromatic aberration of the condenser is troublesome and a condenser similar to the objective is used. Chromatic aberration of the ocular is usually not serious; hence the ocular may be of conventional form, although usually constructed of fused silica, fluorite, or other materials transparent to the infrared or ultraviolet. The ocular may be omitted for special applications in which the primary image formed by the objective is examined directly. [D.S.C.]

Uses. The reflecting microscope is used for examination with nonvisible radiation, or when it is desirable to focus the microscope with visible light and examine with nonvisible ultraviolet radiation. Because reflecting objectives lack chromatic aberration, the focus is the same for different wavelengths. However, in catadioptric objectives with both refracting and reflecting surfaces, spherical aberration occurs.

The reflecting microscope is primarily used in the form of a microspectrophotometer for examination with ultraviolet radiation. Different proteins and nucleic acids have different absorption curves, and the ultraviolet instrument can be used to estimate the amounts of these and other materials. The resolution is better with ultraviolet radiation than with visible light, and this increased resolving power is helpful.

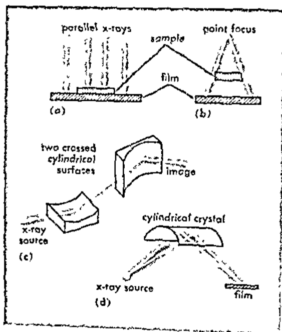
The infrared microscope also requires reflecting optics and can reveal detail in materials that are opaque to light, for example, molybdenum is opaque to light but shows a distinct fiber structure with infrared. Wood, corals, and many red-dyed materials can be examined with the infrared microscope. It is useful in the detection of forgeries. One disadvantage of infrared microscopy in biology comes from the absorption of infrared radiation by water. Dried specimens are usually different in form from living ones. When materials can be dried, or when the water absorption can be subtracted, the infrared microscope gives information concerning the chemical structure of the specimen. Resolving power becomes proportionately less as the longer wavelengths of infrared are used. [O.W.R.]

Microscope, x-ray

A term used to describe a technique and an instrument or combination of instruments which utilize x-radiation for chemical analysis and for magnification of 100–1000 diameters. The resolution possible is about 0.25 micron (μ). X-ray microscopy is a relatively recent development among the microscopic techniques. The contrast in the x-ray microscopic image is caused by the varying x-ray attenuation in the specimen. The advantage of x-ray microscopy is that it yields quantitative chemical information, besides structural information about objects, including those which are opaque to light (*see* X-RAY OPTICS). It is a reliable ultramicrochemical analytical technique by which amounts of elements and weights of samples as small as 10^{-12}–10^{-14} g can be analyzed with an error of only a few per cent.

Principles of image formation. There are four general principles of x-ray microscopy: (1) contact microradiography (Fig. 1*a*), (2) projection x-ray microscopy (Fig. 1*b*), (3) reflection x-ray microscopy (Fig. 1*c*), and (4) x-ray image spectrography (Fig. 1*d*). For a discussion of the first two *see* MICRORADIOGRAPHY.

Contact microradiography. In contact microradiography the thin specimen is placed in close contact with an extremely fine-grained photographic emulsion which has a resolution of more than 1000 lines/mm, and radiographed with x-rays of suitable wavelength. Thus an absorption image in scale 1:1 is obtained and this image is subsequently viewed in a light microscope. The maximal resolution is that of the optical microscope (0.25 μ), but the image has more information which can be ob-



Schematic principles for x-ray microscopy. (a) Contact microradiography; (b) projection x-ray microscopy; (c) reflection x-ray microscopy; (d) x-ray image spectrography.

tained by examining the microradiogram in the electron *microscope*. The optical microscope gives more information because it is possible to correlate the intensity of light in the image with the primary attenuation in the specimen. This is not at present possible in the electron microscope. The latter at present only gives high magnification, but it is not possible to draw conclusions about the quantitative attenuation of electrons in the specimen in order to draw conclusions about the specimen composition. *See* MICROSCOPE, ELECTRON.

Projection x-ray microscopy. Projection x-ray microscopy or x-ray shadow microscopy is based on the possibility of producing an extremely fine x-ray focal spot. This is achieved by an electronic lens system similar to that in the electron microscope. The fine focal spot is produced on a very thin metal foil which serves as a transmission target. The x-rays are generated on the target by the impact of the electrons. The sample is placed near the target and the primary magnification depends on the ratios of the distances from focal spot to sample and sample to film. Resolution is of the same order as the size of the focal spot; at present the best value is about 0.1 μ in favorable objects.

Reflection x-ray microscopy. The method of reflection x-ray microscopy is based on the fact that the refractive index for x-rays in solids is a very small amount less than 1. Thus at grazing incidence (that is, incidence at very small angles), the x-rays will be totally reflected and if the reflecting surface is made cylindrical, there will be a focusing action in one dimension. By crossing two such surfaces a true image formation can be obtained, although with some astigmatism, which can be corrected by giving the surfaces a complicated optical shape. The resolution by this procedure is about 0.5–1 μ.

X-ray image spectrography. X-ray image spectrography utilizes Bragg reflections (*see* X-RAY DIFFRACTION) in a cylindrically bent crystal and produces slightly enlarged emission images; the technique is best classified as a micromodification of x-ray fluorescence analysis. The resolution is about 50 μ. *See* X-RAY FLUORESCENCE ANALYSIS.

Chemistry of specimen. The penetration of x-rays through material varies with the amount and composition of the attenuator (the material) and the wavelength of the x-rays. The specimen must be thin in order to obtain high resolution, and therefore soft (low voltage) and ultrasoft x-rays must be used. For biological specimens x-rays of energie between 0.2 and 5 kilovolts (about 50–2 A in wavelength) are generally used. Thin metallurgical specimens can be examined with harder x-rays because of their higher x-ray absorption. In general the x-ray absorption increases with increasing wavelength of the x-rays (softer rays) and increasing atomic number of the elements composing the specimen. Thus a microscopic structure containing elements of high atomic number of sufficient concentration embedded in a matrix of elements with low atomic numbers will show up as a heavily x-ray-absorbing structure. At certain wavelengths in the x-ray absorption spectrum discontinuities appear: the so-called x-ray absorption edges. *See* X-RAY(S), PHYSICAL NATURE OF. The position of these edges in the spectrum is unique for each element; thus they can be utilized for identification of elements in microscopic structures in a specimen.

Quantitative microscopy. By measuring the variation of density in the x-ray microscopic image the x-ray attenuation can be calculated. If the x-ray microscopic image was recorded with x-rays of suitable wavelength (often monochromatic x-rays), certain chemical characteristics of the specimen can be quantitatively assessed. If a certain element is to be determined in a microstructure, two x-ray microscopic images are recorded with monochromatic x-rays with wavelengths on either side and close to the absorption edge for the particular element. In this way elementary analysis can be performed on specimens weighing not more than 10^{-10}–10^{-12} g with a relatively high degree of accuracy. Instead of using a photographic film to record the x-ray image, various types of detectors (Geiger-Müller tubes) are used to measure the variations of x-ray transmission in the sample. Such techniques, especially in the form of scanning, may become more useful in the future.

By proper selection of x-ray wavelength the dry weight, water content and the content of certain other compounds can be determined in cellular structures down to about 1 μ in size. Thus weights as small as 10^{-14} g can be determined with an analytical error of only a few per cent.

Applications. In biology, x-ray microscopy has been utilized for the quantitative determination of the dry weight, water content and elementary composition of many tissues, for example, nerve cells, various types of secretory cells, the individual bands in muscle fibers, chromosomes, and parts thereof. Especially in studying mineralized tissues much new information has been gained through the use of x-ray microscopy.

The capillary circulation in the living animal can be studied by x-ray microscopy. Thorotrast and Umbradrie, contrast media, may be injected into the bloodstream and successive microradiograms recorded. The procedure is called microangiography and has been applied to the study of the finest blood vessels.

As indicated previously, soft and ultrasoft x-rays must be used for the x-ray microscopy of thin sections or smears of biologic tissue. These long-wavelength x-rays must be generated in specially designed x-ray tubes. The sample and photographic emulsion are enclosed within the high vacuum of the x-ray tube. The pictures obtained show the distribution of dry weight (mass) within the cells and tissues in addition to information on the structures in the specimen.

It is possible to determine the thickness of thick samples by using oblique incidence of the x-rays. With this technique, the thickness of nerve fibers and some constituents of bone tissue has been determined.

By tilting the film and specimen at a certain angle, stereoscopic microradiograms have been obtained. When thick sections of tissue have been used, the 3-dimensional arrangement of bone cells in bone tissue and the 3-dimensional image of the capillary net in the circulatory system have been made available for study.

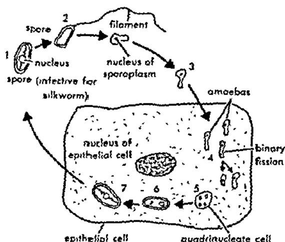
X-ray microscopy has also been applied to inorganic and organic chemistry, mineralogy, and metallurgy. [A.EN.]

Bibliography: V. E. Cosslett, A. Engström, and H. H. Pattee, Jr. (eds.), *X-ray Microscopy and Microradiography*, 1957.

Microsporidia

An order of Cnidosporidia characterized by the production of minute spores with a single intrasporal or 1 or 2 intracapsular filaments and a single sporoplasm. The spore membrane is usually a single piece. Microsporidians are mainly intracellular parasites of arthropods and fishes.

The sporoplasm or amebula, released from the spore in the intestine of the host, passes through the gut wall to reach the site of infection by way of the blood stream. The amebula enters a cell and becomes a trophozoite, feeding and growing at the expense of the host. Eventually, the trophozoite divides by binary fission, or schizogony. The cells resulting from this asexual division develop into sporonts. A sporont may transform directly into a sporoblast and give rise to a single spore, or it may



Diagrammatic life cycle of Nosema bombycis, the cause of pébrine, a fatal disease of silkworms. (1) The spore is typical of the order, with a single filament and sporoplasm. Upon ingestion, the filament is extruded (2) and the sporoplasm enters an epithelial cell of the gut (3) where it divides many times to form uninucleate amebas which fill the cell. The cell may become dissolved. (4) Then the amebas attack other cells, perhaps almost all of the cells in the body of the worm. Eventually the parasites form 4-nucleate cells (5), and these develop into the spore (6, 7). Silkworms become infected by eating mulberry leaves or other food contaminated by feces or body fragments of infected worms.

undergo further nuclear division and produce several sporoblasts, each of which will form a spore. In this order there is no evidence that special cells differentiate in the developing spore to form valves or a capsule as occurs in the Myxosporidia and Actinomyxidia.

An interesting feature of microsporidiosis is that the parasites may induce an extreme enlargement (hypertrophy) of the cell or nucleus. Thus, certain cells of the stickleback and smelt, infected with *Glugea anomala* and *Glugea hertwigi* respectively, often enlarge from a diameter of 8–10 microns (μ) to one of 5000 μ. Such cells are called *Glugea* cysts.

The microsporidian *Nosema bombycis* is a parasite of the silkworm, the larval stage of the insect *Bombyx mori* L. Invasion by the parasite of all the cells of the larva, pupa, and adult, causes the lethal disease pébrine which was first studied in detail by Pasteur in 1865. *See* CNIDOSPORIDIA; *see also* LEPIDOPTERA. [R.F.N.]

Microtatobiotes

A class of microorganisms of the division Protophyta of the plant kingdom. Members are the smallest of the known living things and cause diseases of plants, animals, and man. Individuals range in size from typhuslike organisms (rickettsiae) visible under a light microscope, down to filterable viruses

fluid, such as blood, is placed on the end of a clean slide; a second slide is touched to it (Fig. 4) and pushed forward to leave a uniform film.

Blood smears of this type are usually stained in a methylene blue-eosinate dissolved in methanol, of which Wright's stain is a typical example. This stain also requires a phosphate buffer with pH of 6.4.

The air-dried smear is flooded with a specific amount of stain and laid on a rack, or across the top of a beaker, for 1 min. Buffer is then added to the stain from a drop bottle in the proportion of two drops of buffer to each drop of stain. After 2 min, the mixture is washed from the slide with distilled water. Blood smears are usually preserved in the dry state.

Squashes are just what their name indicates, and their success is dependent on the condition of the material selected. Anthers of plants, or the testes of insects, may be squashed directly, but root tips or plant ovaries require softening before squashing. Cellulases derived from the snail's stomach are much used for this preliminary.

The salivary glands of *Drosophila* are so widely prepared by this technique that they will serve as an example. Glands are taken from a third instar larva, easily recognized by the sluggish movements

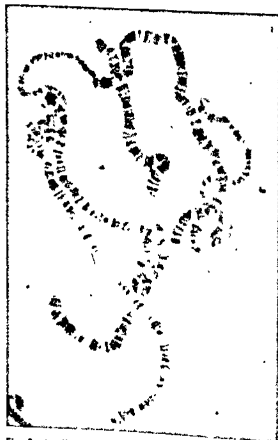


Fig. 5. Well-spread chromosomes shown with the bands clearly marked. (From P. Gray, *Handbook of Basic Microtechnique*, 2d ed., McGraw-Hill, 1958)

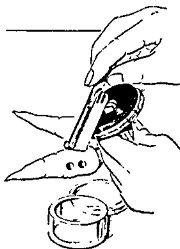


Fig. 6. Cutting hand section. razor is across the with gentle sure and the tion into a s (From P. G *Handbook of Microtechnique* ed., McGraw-Hill, 1958)

with which it crawls on the side of the culture vessel, by pulling them out with the head. This is best performed in a drop of LaCour's acetoorcein Excess tissue is rapidly stripped away, a coverslip placed on top, and strong pressure exerted with the forefinger through a sheet of bibulous paper. Inexperienced technicians may find it desirable to dissect out the glands in a drop of saline before transferring them to stain. A successful preparation (Fig. 5) shows sharply stained chromosomes against a faintly pink background.

These preparations may be rendered permanent. The slide is first frozen for about 5 min on a block of dry ice and the coverslip then gently levered off with a safety razor blade. The squashed material adheres to the slide which is then placed in a coplin jar containing 70% ethanol before being dehydrated, cleared, and mounted in the same manner as a section.

Sections. All investigations of the structure of cells, or of the relations of cells within organs, require the examination of thin slices known biologically as sections. The useful range of thickness for examination under the optical microscope is from about 25 to 2 μ; 10-μ sections are routine in most histological and pathological work.

Thick sections of crisp or waxy structures, particularly plant stems and roots, may be cut by hand (Fig. 6). Such sections are subsequently stained and mounted, as though they were wholemounts, by the methods described above.

Most biological specimens, however, do not have the necessary consistency for such treatment be- must be impregnated with, and embedded in, substance which sections well. Paraffin is the widely used. The preparation of paraffin secti requires the following steps: fixation, preservati dehydration, clearing, impregnation, embeddi cutting, mounting on slides, dewaxing, rehydrati staining, differentiating, counterstaining, drating, clearing, and mounting.

Fixation, preservation, and dehydration of pieces of tissue are the same as for the whol described above. Clearing, however, is in this







(a) Section of pituitary gland, Mallory-Azan stain.

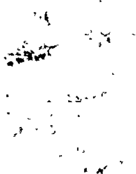
(b) Section of kidney; Feulgen-fast green stain.

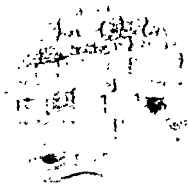
(c) Section of kidney through a glomerulus; Mallory triple stain.

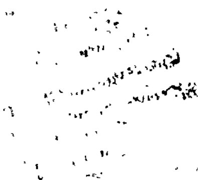
(d) Intestinal villus; hematoxylin and eosin.

(e) Intestine; silver impregnation, azocarmine and light green.

(f) Onion root tip showing stages of mitosis, Flemming triple stain (Allied Chemical and Dye Corp)







preliminary to impregnation with molten wax, and xylol is the reagent of choice. Five-millimeter cubes of animal tissue should have at least 2 hours in one change of 70% ethanol, two changes of 95% ethanol, two changes of pure ethanol, and two changes of xylol before being placed in molten wax. Plant tissues, in which the protoplasm tends to shrink from the cell wall, should be passed through the following sequence of reagents: (1) 5% ethanol, (2) 10% ethanol, (3) 18% ethanol, (4) 30% ethanol, (5) 40% ethanol plus 10% *tert*-butanol, (6) 50% ethanol plus 20% *tert*-butanol, (7) 50% ethanol plus 35% *tert*-butanol, (8) 40% ethanol plus 55% *tert*-butanol, (9) 25% ethanol, 75% *tert*-butanol, (10) pure *tert*-butanol. It should be understood that, for example, solution (7) of this series is prepared by mixing 50 parts of ethanol with 35 parts of butanol and then "making up" to 100 parts, by addition of water. Five-millimeter cubes of plant tissue will be adequately dehydrated and cleared after 1 hour in each of the solutions except (6), in which the pieces should remain for 6–12 hours.

The tissues, whether cleared in xylol or butanol, must now be transferred to molten paraffin, maintained at just above its melting point in a thermostatically controlled oven or bath. Wax of 54–56°C melting point, held at 58°C, is conventional. Softer wax, 52–54°C, is difficult to cut but should be used, at 55°C, for muscular tissues, such as heart and tongue, which tend to harden at 58°C. Harder waxes are rarely used. It is customary to add wax shavings to the tube containing delicate specimens in solvent, to leave this overnight, and then to place the tube in the oven. Pieces of homogeneous tissue like liver may be transferred directly to molten wax.

The impregnated pieces must now be embedded in a wax block. To this end paper boats, or other suitable containers, are filled (Fig. 7) with molten wax, the specimen is placed in the boat, a heated pipette is used to melt any film which may have

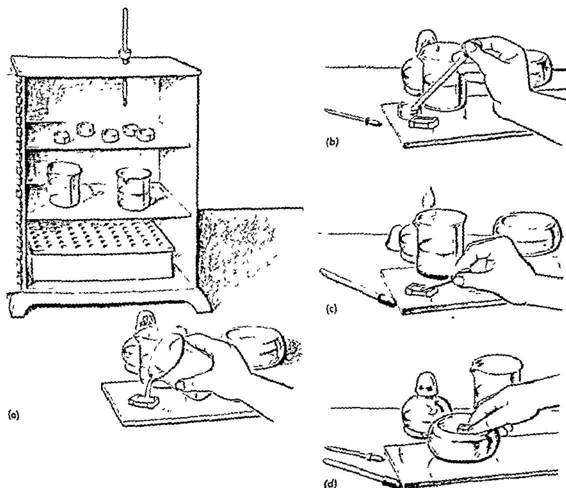


Fig. 7. (a) Filling with wax an embedding box which has been attached with water to a glass slide. (b) Transferring the object from the embedding dish to the wax-filled paper box. (c) Remelting the wax around the object with a heated pipette. (d) Cooling the wax block. (From P. Gray, *Handbook of Basic Microtechnique*, 2d ed., McGraw-Hill, 1958)

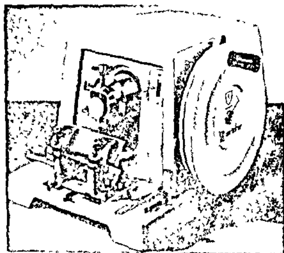


Fig. 8. American Optical Company rotary microtome. (From P. Gray, *Handbook of Basic Microtechnique*, 2d ed., McGraw-Hill, 1958)

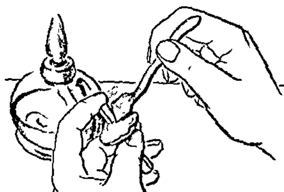


Fig. 9. Mounting the wax block on the block holder. (From P. Gray, *Handbook of Basic Microtechnique*, 2d ed., McGraw-Hill, 1958)

formed around the specimen, and the block is then chilled and solidified by partial immersion in water. A satisfactory block is translucent and almost flat on top. A chalky block has either been cooled too slowly or made from solvent-contaminated wax. A block with a deep conical depression in the top has been cooled too fast.

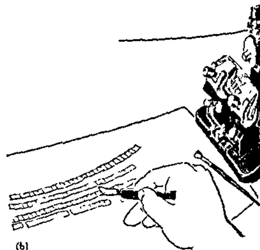
Sections are cut from the block after it has been mounted on a microtome (Fig. 8). The revolution of the handle of this device raises and lowers the block vertically against a knife. The block is advanced horizontally at the top of its rise by the action of a micrometer screw advancing a pressure point along a spring-loaded diagonal plate. The extent of the advance, and thus the thickness of the section, is controlled by varying the moment of engagement of the pawl with the ratchet wheel attached to the micrometer screw.

The block of wax containing the object is removed from the container in which it was cast and trimmed until about equal volumes of specimen and wax remain. It is essential that two of the faces be parallel. The trimmed block is fused (Fig. 9) to the object carrier of the microtome. The carrier is placed in the holder of the microtome with the parallel faces parallel to the knife edge, adjusted until it almost touches the knife, and the handle turned until the ribbon of sections starts to form. The commonest defects are that the ribbon splits because of a nick in the knife edge or that the sections roll into cylinders without forming sections. This last is due to a faulty relationship between the hardness of the block and the angle at which the knife strikes it. This relation must be established empirically for each block. The block lifting the ribbon from the knife blade is sometimes due to a faulty angle and sometimes to a dirty knife.

When enough sections have been cut, the ribbon is divided (Fig. 10) into pieces about 2 in. long in preparation for mounting. A clean slide is lightly smeared with a mixture of equal parts egg albumen



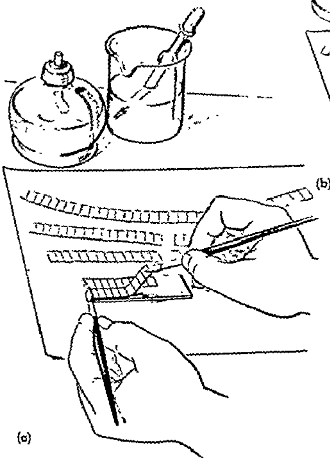
(a)

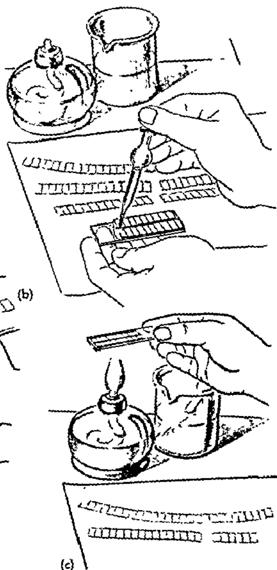


(b)

Fig. 10. (a) Starting the paraffin ribbon. (b) Cutting the ribbon in lengths. (From P. Gray, *Handbook of Basic Microtechnique*, 2d ed., McGraw-Hill, 1958)

Fig. 11. (a) Mounting the dry ribbon. (b) Flooding the ribbons. (c) Warming the flooded ribbons to flatten them. (*From P. Gray, Handbook of Basic Microtechnique, 2d ed., McGraw-Hill, 1958*)





and glycerin. The pieces of ribbon are laid on the prepared slide and flooded with water before being gently warmed (Fig. 11) until flat. The excess water is then drained off and the slide set on a warming table, at about 35°C, until dry.

The sections are now ready to be stained. To prepare the conventional celestin blue B–eosin histological slide, the following coplin jars (Fig. 12) should be set up: (1) xylol, (2) 50% xylol-ethanol, (3) pure ethanol, (4) 95% ethanol, (5) 70% ethanol, (6) water, (7) Gray's celestin blue B, (8) (placed behind the 95% ethanol) 0.2% ethyl eosin in 95% ethanol. The slide is warmed until the wax melts, dropped into xylol for 1 min and then passed down the series, with about 30 sec in each jar, until it is in water (6). It is then transferred to stain (7) for from 1 to 2 min, rinsed in water, and transferred up the series again until it reaches 95% ethanol (4). The slide is now dipped up and down in the eosin solution (8) until the sections are sufficiently yellow—a point on which opinions vary widely. A quick rinse in 95% ethanol (4) precedes transfer to pure ethanol (3) for about 30 sec and thence into xylol. The

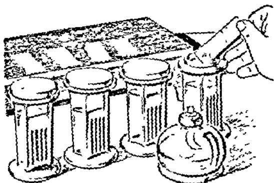


Fig. 12. Starting a slide through the reagent series. (*From P. Gray, Handbook of Basic Microtechnique, 2d ed., McGraw-Hill, 1958*)

slide is then withdrawn from xylol, an adequate amount of balsam is placed on the surface and a coverslip applied.

Finishing and storing slides. Mounts made with balsam should be hardened on a warm plate at about 45°C for 3 or 4 days. Surplus exuded balsam may then be wiped off with a rag moistened with xylol. Another day of hardening will prepare the slide for a final cleaning in warm soapy water, after which the permanent label may be attached.

Wholemounts must be, and sections are better, stored in flat trays than in vertical grooves. Slides so stored, in a cool dark place, do not alter appreciably in 50 years. [P.GR.]

Bibliography: J. R. Baker, *Discovery of the Uses of Coloring Agents in Biological Microtechnique*, 1945; H. J. Conn, *Biological Stains*, 6th ed., 1953; P. Gray and F. Gray, *Bibliography of Biological Microtechnique*, 1956; P. Gray, *The Microtomist's Formulary and Guide*, 1954; J. E. Sass, *Botanical Microtechnique*, 2d ed., 1951.

Microwave

An electromagnetic wave which has a wavelength in the centimeter range. Microwaves occupy a region in the electromagnetic spectrum which is bounded by radio waves on the side of longer wavelengths and by infrared waves on the side of shorter wavelengths (see Fig. 1). There are no sharp boundaries between these regions except by arbitrary definition. In the microwave region, a further delineation is sometimes made with such names as decimeter, centimeter, or millimeter waves.

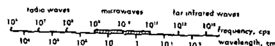


Fig. 1. A portion of the electromagnetic spectrum.

The historical development of microwaves is only a particular phase of the gradual evolution of the concept and application of electromagnetic waves in general. The foundation for the entire field was laid by James Clerk Maxwell in 1864 when he formulated a set of equations governing electromagnetic phenomena which became known as Maxwell's equations. A solution of these equations led Maxwell to predict the existence of electromagnetic waves when none were known to exist at that time. The next decisive development came when Heinrich Hertz demonstrated in 1888 an experimental proof for the existence of electromagnetic waves and verified substantially all aspects of Maxwell's predictions. Once the validity of the basic theory was established, further generalizations and applications became a matter of technical development. *See* MAXWELL'S EQUATIONS.

In his classical experiments, Hertz used damped electromagnetic waves in the decimeter or meter range. It might seem that this demonstration was the starting point for modern microwave techniques. Actually, the detailed development of microwaves lagged many years behind that of radio waves and got its start only in the 1930s when continuous-wave microwave generators were invented and hollow wave guides and cavity resonators were introduced. The drastic change occurred not because of the selection of any particular portion of the electromagnetic spectrum but in the exercise of a new approach which was particularly amenable to the microwave range. This new approach resulted from an application of the solutions of Maxwell's equations for apparatus dimensions which are comparable to the microwave wavelength.

Frequency and velocity. Although an electromagnetic wave is commonly referred to in terms of its wavelength, a more fundamental characteristic of a wave lies in its frequency. As in any type of wave motion, these two quantities are related to a third quantity, velocity of propagation, by the simple relation

$$f\lambda = v$$

where f = frequency, λ = wavelength, and v = velocity of propagation. *See* WAVE MOTION.

It is usually simpler and more definitive to refer to a wave by its frequency than by its wavelength. The only exception is in the case of wave propagation in free space (in vacuum or, approximately, in air) where a wave propagates with the velocity of light (approximately 3×10^{10} cm/sec); the wavelength is then a perfectly definitive quantity. According to usual practice, when a wavelength is mentioned without qualification, it means the free-space wavelength. Thus, for a microwave wavelength of 3 cm, the corresponding frequency is 10^{10} cycles per second.

Generation of microwaves. Microwaves can be generated as direct radiation from electrical sparks across gaps by applying a high electric potential. The spark gap can also be a part of a very high-frequency oscillating circuit which radiates electromagnetic waves. Microwaves can also be derived from the thermal radiation of warm bodies. But all these sources are unsatisfactory because of the lack of purity of the wave and the low power of the radiation. In contrast to these, all modern microwave generators are electronic devices which produce continuous-wave (CW) oscillations of a single tunable frequency. Some important microwave generators are known as klystrons, magnetrons, and traveling-wave oscillators. Their power outputs range from microwatts to thousands of kilowatts, depending upon the type and design of the generator and the operating frequency. For details *see* KLYSTRON; MAGNETRON; MICROWAVE TUBE; TRAVELING-WAVE TUBE.

Microwave circuit elements. Any particular grouping of physical elements which are arranged or connected together to produce certain desired effects on the behavior of microwaves is known as a microwave circuit. It should be noted that a microwave circuit or any of its elements is not "closed" in the sense of a low-frequency electric circuit. Figure 2 shows a microwave circuit in

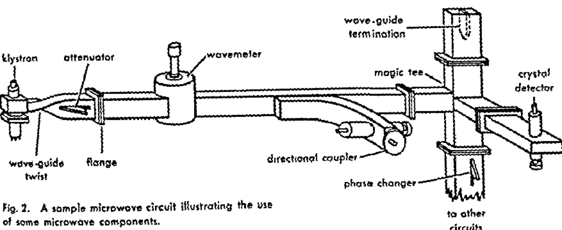


Fig. 2. A sample microwave circuit illustrating the use of some microwave components.

which some circuit elements to be described later are illustrated.

Microwave wave guides. A wave guide is a circuit element which constrains or guides the propagation of microwaves along a path defined by the physical construction of the guiding element It can be, for instance, a coaxial cable having an outer conductor of annular cross section coaxially placed with respect to an inner conductor By prevailing usage, however, a microwave wave guide usually means a hollow metallic tube which can confine and guide the propagation of microwaves

When an electromagnetic wave propagates in a hollow wave guide, the electric and magnetic intensities (denoted by E and H, respectively) must satisfy both Maxwell's equations and certain boundary conditions. The result is that only certain specific patterns for the distribution of E and H (taken together) can exist in the wave guide. Each unique pattern of field distribution is called a mode. There are two types of mode possible in a hollow wave guide—the transverse electric (TE) mode, and the transverse magnetic (TM) mode. For additional discussion, *see* WAVE GUIDE.

One of the most important characteristics of a TE or TM wave is that it has a cutoff wavelength for each mode of transmission. If the free-space wavelength is longer than the cutoff value, that particular mode cannot exist in the wave guide. For any given wave guide, the mode that has the longest cutoff wavelength is known as the dominant mode. Figure 3 illustrates the dominant mode TE_{01} for a rectangular wave guide.

Rectangular wave guides having cross sections uniform through their entire length are by far the commonest in use. Wave guide sections may be straight, bent or twisted. Several wave guides may form a network such as a T junction, a magic tee, or a directional coupler (*see* DIRECTIONAL COUPLER).

Attenuators. A microwave attenuator is a device that causes the field intensity of a wave to decrease by absorbing a part of the incident power. This objective is usually accomplished by inserting a piece of lossy material in the wave guide along the direction of electric intensity. The lossy material can be simply a sheet of material containing powdered carbon whose electrical conductivity is in the proper range to cause the desired attenuation. For an adjustable amount of attenuation, it is usually arranged that the absorbing sheet can be mechanically moved by varying degrees in or out of the path of wave propagation.

Phase changers. A phase changer is a device that causes the field intensity to shift its phase without attenuating its amplitude. This can be done by inserting a thin slab of low-loss dielectric material parallel to the direction of electric intensity. Or, for small phase shifts, a metal pin in the form of a screw can be used to penetrate variable amounts into the wave guide. Still another way of changing the phase of a wave is to vary the path length

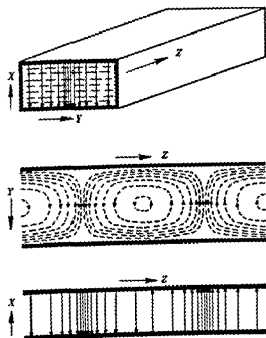


Fig 3. Instantaneous field pattern for the TE_{01} wave in a rectangular wave guide. The wave propagates in the z direction. Solid lines indicate the electric intensity E and dashed lines the magnetic intensity H.

of a reflected wave by moving a short-circuiting plunger, which is a metal block fitted snugly in the wave guide to reflect totally all incident waves.

Detectors. A microwave detector is a device that can demonstrate the presence of a microwave by a specific effect that the wave produces. One of the most effective means of detection makes use of a nonlinear device which converts the microwave field intensity into either a direct current or a low-frequency alternating current. A silicon crystal making a pin-point contact with a tungsten wire is perhaps the most commonly used detector for microwaves. This semiconducting material has the nonlinear characteristic that the resistance for a potential of one polarity is drastically different from that for the opposite polarity. Hence, an alternating potential produces unequal currents in the positive and negative portions of the cycle, giving an average current which can be indicated by a direct-current (dc) meter. A typical crystal detector gives a current of the order of 1 milliampere for a microwave power input of 1 milliwatt. One important reason why a crystal detector is particularly adaptable to the detection of high-frequency waves, including microwaves, is the exceedingly small capacity in the contact area between the tungsten point and the crystal. The contrary case of a larger capacity would cause the high-frequency currents to bypass the detector, which would then become ineffective.

A crystal detector for microwaves is usually placed in a special wave-guide mount which consists of a wire post positioned along the direction of strongest electric intensity in the wave guide. *See* DETECTOR.

Another type of microwave detector in common use is called a bolometer. It is a device whose resistance changes sensitively with temperature, the latter being a function of the absorbed microwave power. The bolometer can be used as one arm of a resistance bridge circuit in which any change in resistance causes an unbalance in potential or current, which can be measured by a suitable meter. *See* BOLOMETER.

Probes. Either of the aforementioned microwave detectors can be used in a wave guide where the detector is designed to absorb almost all the power that impinges on it. In this case, the detector is said to form a matched termination for the wave guide. At other times, however, the object of the operation is to probe into the character of the field intensity at any point with a minimum disturbance of the field distribution. Under these conditions, a detector is used in conjunction with a small pin which is inserted into the field by the least amount that makes detection possible. Such a detector is called a probe. It is used primarily for measuring the standing-wave ratio, which gives a measure of the interference pattern formed by the incident and reflected waves.

Wavemeters. A microwave wavemeter is usually made of a tunable resonant cavity which measures the free-space wavelength (or frequency) of a wave by the position of tuning at resonance (*see* CAVITY RESONATOR). The resonance frequency relative to the tuning position can be calculated accurately from the known dimensions of the cavity or calibrated against a known frequency standard.

Transmission of microwaves. A microwave transmitter is similar in all its principal aspects to an ordinary radio transmitter. It consists of a microwave generator, a power amplifier (if necessary), a circuit containing all necessary elements, means of modulation to impart some form of information or program to the waves, and an antenna network to send the waves out into space. Microwave generators and various circuit elements have been mentioned in the preceding sections. Since World War II, klystron amplifiers have come into practical use at high powers and traveling-wave amplifiers at low powers.

Modulation processes. A microwave can be transmitted as a continuous wave without any form of modulation. As a means of communication, a CW transmission is useful only in coded messages such as dots and dashes when the transmission is keyed on and off. To enrich the information content of a transmission, methods involving amplitude modulation (AM), frequency modulation (FM), or phase modulation (PM) can be used. In amplitude modulation, the amplitude of the wave is varied at some rate (audio or video) while the frequency is kept constant. In frequency modulation, the frequency of the wave is varied at some rate while the amplitude is kept constant. Phase modulation is so similar to frequency modulation in many respects that a distinction between the two will not be drawn here. Generally speaking, the net result of each type of modulation is to form two sidebands (each sideband containing instantaneously more than one side frequency for frequency modulation) that are symmetrically located around the central carrier frequency, which is provided by the generator. Thus, the whole transmission occupies a channel width twice as wide as each sideband. *See* AMPLITUDE MODULATION; FREQUENCY MODULATION; PHASE MODULATION.

Pulse operation of microwave transmitters is a special modulation process which is important enough to deserve particular mention. Transmission by pulsed waves is an efficient method of sending an extraordinarily large peak power at the expense of a small average power. For example, if a microwave generator is operated at the rate of 500 pulses per sec and each pulse lasts only for 1 microsecond, then it can deliver a peak power of 200 kilowatts at the expense of only 100 watts. But, even apart from the power considerations, pulse transmission is the most effective way of securing echo signals by reflection, as in radar. *See* PULSE MODULATION.

Microwave antennas. A transmitting antenna is a circuit element which transfers the wave excitation from the generator into waves propagating in

free space. There are two major considerations in the design of an antenna system: sensitivity and directivity. Sensitivity of a transmitting antenna is measured in terms of the ratio of the power delivered into space to the available excitation power, and naturally should be made as high as possible. Directivity refers to the pattern of the field intensity distribution in all directions at a fixed radial distance from the antenna. Although the degree of directivity depends upon the specific requirements of a given situation, highly directive antennas are necessary in most microwave applications. Figure 4 illustrates a simple microwave antenna using a parabolic reflector. For more detailed information, *see* ANTENNA (AERIAL); DIRECTIVITY; MICROWAVE TRANSMISSION LINES.

Microwave propagation in space. The simplest as well as the most important fact about electromagnetic wave propagation in space is that a wave tends to travel along a straight line in the absence of obstructions. In the application of microwaves, straight line, or line of sight, propagation in the atmosphere is often implicitly assumed. However, if the transmitted beam either wholly or partially hits the earth's surface or any physical obstacle, it will be reflected, with a possible loss of energy, depending on the shape, size, and the reflective properties of the obstructing medium. In fact, it would be even more correct to say that the wave is in general scattered by various objects with the particular phenomenon variously described as reflection, refraction, or diffraction. (For an explanation and discussion of these phenomena, *see* MICROWAVE OPTICS.) A similar situation arises when a microwave has a "sky" component which will be little affected by the atmosphere in the troposphere and stratosphere regions, but will be more strongly influenced by the ionosphere. Unlike radio waves of lower frequencies, which are often reflected by the ionosphere, a microwave has too high a frequency to be reflected by this medium at angles near vertical incidence. However, for microwaves slanted toward the horizon, reflection by the ionosphere becomes possible. *See* RADIO-WAVE PROPAGATION.

Although the attenuation of a microwave by the earth's atmosphere is usually small, it may be selectively large for certain bands of microwaves. For instance, the absorption by water vapor is selectively high for waves in the K-band (wavelength around 1 cm) and that by oxygen is high for waves in the so-called $\frac{K}{2}$-band (wavelength around ½ cm). These considerations are important in the choice of wave bands for the purpose of microwave communication.

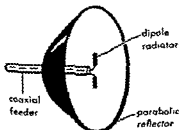


Fig. 4. A microwave antenna consisting of a parabolic reflector, coaxial feeder, and dipole radiator.

Reception of microwaves. The microwave-transmitting antennas mentioned earlier can be used for reception, where the direction of wave propagation is in the reverse order. Since the received wave is often appreciably weaker than the transmitted wave, reliable reception depends upon the sensitivity of the receiver in picking up weak signals in the midst of noises. The receiver sensitivity is specified by the minimum detectable voltage at the input end of the receiver. A typically good receiver has a sensitivity of about 1 microvolt. In most cases, such a receiver makes use of a superheterodyne scheme which involves a local microwave oscillator beating with the incoming wave to form a so-called intermediate frequency (i-f) signal for amplification and further detection. A crystal is generally used for the mixer, while automatic frequency control (AFC) is often used to keep the i-f constant in frequency. The bandwidth for the amplifier should be wide enough to pass both sidebands of the signal, but not much wider than necessary in order to minimize the noise.

Radar. Radar is the abbreviated name for radio detection and ranging. This apparatus is used to detect the presence and the range of a reflecting object by transmitting a pulsed microwave and receiving the reflected wave of the same pulse. Also, the direction of the reflecting object can be determined from the orientation of the directive antenna at maximum reception intensity. Initial development of radar began secretly around 1932, but active use of radar first occurred around 1940 during World War II. *See* RADAR.

There are many types of oscilloscope presentations used for various purposes in radar. There is in particular a special presentation known as plan-position indicator (PPI), in which the echo strength is made to intensify a spot on the oscilloscope screen while the position of the spot corresponds to specific orientations of the scanning antenna. When used in an airborne radar, PPI presentation gives a panoramic view of the surface structures of the ground.

Advantages of microwaves. From the point of view of radio communication, a great advantage of microwaves lies in the immense spaciousness of useful frequencies. For instance, the frequency difference between the S-band (wavelength around 10 cm) and K-band (wavelength around 1 cm) is roughly 20,000 Mc/sec. This frequency "space" is about 100 times the combined frequency range of present-day radio broadcasting, communication, and television (*see* RADIO SPECTRUM ALLOCATIONS). There is a good prospect that this immense range can be increased 5–10 times by further improv

ment of power generation at the high-frequency end of the microwave range. Another advantage of microwaves is related to the high directivity and resolving power of microwave wavelengths. Narrow microwave beams can be readily formed by antennas of physically convenient sizes. The superiority of microwaves over the waves of much longer wavelengths, with respect to radiation directivity, is made evident by the very large antenna structures necessary for beam transmission at longer wavelengths. Moreover, the resolving power, as measured by the ability of a wave to differentiate one reflecting object from another, is larger for microwaves than for longer waves.

The reader may wonder at this point, if the advantages of microwaves are attributed to properties of very short waves, why is it that shorter waves like infrared radiation would not be even superior for various applications. The answer is that infrared radiation is being used in many applications where its advantages are well proved. However, there are some major differences between these regions of radiation. Unlike microwaves, infrared radiation is usually nonmonochromatic in frequency, noncoherent in phase, and not easily subject to manipulations like frequency modulation, amplification, and electronic control. Also, infrared radiation probably cannot compare favorably with microwaves with regard to power concentration at a given frequency and range of propagation in the atmosphere. These factors tend to make microwaves a more useful tool for communication purposes. *See* INFRARED RADIATION.

Applications of microwaves. The most important practical application of microwaves is radar. In fact, microwave techniques have been developed mainly under the incentive of radar work. Next to radar, the use of microwaves as carrier waves in relay links for multichannel transmission of telephone, telegraph, and television has already become practically important.

In postwar years, application of microwaves has flourished in many scientific and basic research fields. Microwave spectroscopy has become an established science chiefly because of the availability of microwave instruments and techniques. The knowledge of the structure of numerous molecules and crystals has been greatly increased by this new branch of spectroscopy. To mention a celebrated case, the fine structure of the hydrogen atom known as the Lamb shift was discovered by the use of microwaves. Another important advance came with the development of the so-called atomic clocks, which use microwave resonance interactions with either cesium atoms or ammonia molecules. The introduction of the ammonia maser led to the development of the solid-state maser, which is virtually a noiseless amplifier. One of the most important applications of the maser is in the field of radio astronomy, which is in its own right a fertile ground for microwave application. For instance, the hyperfine line of the hydrogen atom at 1420 Mc/sec has been observed in stellar radiation with refined microwave techniques. In the field of nuclear physics instrumentation, microwave electronics has been applied in a most significant manner to high-energy linear accelerators and similar devices. *See* PARTICLE ACCELERATOR; *see also* ATOMIC CLOCK; MASER; MICROWAVE SPECTROSCOPY; RADIO ASTRONOMY. [C.K.J.]

Bibliography: MIT Radar School Staff, J. F. Reintjes and G. T. Coate (eds.), *Principles of Radar*, 3d ed., 1952; C. G. Montgomery, R. H. Dicke and E. M. Purcell, *Principles of Microwave Circuits*, MIT Radiation Laboratory Series, vol. 8, 1948; H. J. Reich, P. F. Ordung, H. L. Krauss, and J. G. Skalnik, *Microwave Theory and Techniques*, 1953; R. I. Sarbacher and W. A. Edson, *Hyper and Ultrahigh Frequency Engineering*, 1943.

Microwave optics

The study of the properties of microwaves which are analogous to those of light waves in optics. The fact that microwaves and light waves are both electromagnetic waves, the major difference being that of frequency, already suggests that their properties should be alike in many respects. But the reason microwaves behave more like light waves than, for instance, very low-frequency waves for electrical power (50 or 60 cycles per second) is primarily that the microwave wavelengths are usually comparable to or smaller than the ordinary physical dimensions of objects interacting with the waves.

In his classical experiments to verify Maxwell's theory, H. Hertz first demonstrated the optical properties of damped decimeter waves, such as rectilinear propagation, reflection, refraction, and polarization. Today it is almost taken for granted that microwaves inherently possess all these properties, and the language of geometrical or physical optics is freely used wherever allowed by the situation.

Rectilinear propagation. As is the case with light, a beam of microwaves propagates along a straight line in a perfectly homogeneous infinite medium. This phenomenon follows directly from a general solution of the wave equation in which the direction of a wave normal does not change in a homogeneous medium (*see* WAVE EQUATION). In the use of radar, a microwave beam is justifiably presumed to travel in a straight line before and after reflection by an object. For microwave communication in cases when two distant stations are not along the line of sight, the waves would be blocked by the earth's surface. The difficulty is remedied by the use of microwave relay links so that straight line propagation is maintained in each section. *See* MICROWAVE TRANSMISSION LINES.

Reflection and refraction. Consider a plane boundary between two semi-infinite media having different physical properties (Fig. 1). If a plane-polarized microwave is incident from medium I, the boundary conditions generally require the pre-

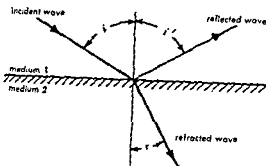


Fig. 1. Reflection and refraction of microwaves at a plane boundary between two insulating media.

ence of a reflected wave back to medium 1 and a transmitted (refracted) wave into medium 2. In the case of two insulating media, the familiar relations in optics hold:

$$i = i' \qquad \text{for reflection} \qquad (1)$$

$$\frac{\sin i}{\sin r} = \sqrt{\frac{\epsilon_2}{\epsilon_1}} = N \qquad \text{for refraction} \qquad (2)$$

where the angles i, i', and r are as indicated in Fig. 1, ϵ_1 and ϵ_2 are the dielectric constants of media 1 and 2 respectively and N is the index of refraction of medium 2 relative to medium 1. The reflected and refracted intensities depend upon whether the incident electric intensity is polarized in the plane of incidence or perpendicular to it. In any case, the well-known Fresnel equations of optics can all be applied to this case. See REFLECTION (ELECTROMAGNETIC RADIATION); REFRACTION OF WAVES.

With some modification the laws of reflection and refraction can be applied to the propagation of microwaves inside a dielectric-filled metallic wave guide (*see* WAVE GUIDE). The usual case is that of a vertical incidence to the plane boundary between two dielectric media perpendicular to the lengthwise direction of the wave guide. The reflection coefficient can be obtained by measuring the standing-wave ratio in the wave guide. Another interesting application is associated with the microwave analog of total internal reflection in optics It may be seen from Eq. (2) that if $\epsilon_1 > \epsilon_2$ (that is, if the wave is incident from a denser medium) there is a total internal reflection for the wave when $i > \sin^{-1}\sqrt{\epsilon_2/\epsilon_1}$. This means that a properly designed dielectric rod (without metal walls) can serve as a wave guide by totally reflecting the elementary plane waves Still another case of interest is that of a microwave lens. By using either a natural dielectric of a certain shape or an artificial dielectric consisting of an array of thin metal plates of a certain design, a microwave lens can be constructed which has the required index of refraction. Such lenses have been used as microwave antennas. For further information, *see* ANTENNA (AERIAL).

The reflection of a microwave by a conducting plane has all the characteristics of the reflection of

... The wave in the conducting medium does not go much beyond a "skin depth" and is of little practical consequence. Examples of reflection of microwaves by conductors are found in parabolic reflectors used as antennas and in targets for radar beams.

Diffraction. In an analogous manner to light, a microwave undergoes diffraction when it encounters an obstacle or an opening which is comparable to or somewhat smaller than its wavelength. Diffraction problems have been

ture but th

for a simp

importance, however, may be cited as an illustration Let two wave guides be coupled through a small hole in a metallic partition as shown in Fig. 2. The radius of the hole is assumed to be much smaller than $\lambda/2\pi$, where λ is the wavelength. A wave in one wave guide will leak through the hole by diffraction into the other wave guide. While an exact calculation is difficult, a satisfactory treatment can be worked out by regarding the diffraction effect as being equivalent to the presence of an electric and magnetic dipole placed at the position of the hole. The radiation fields of the dipoles supply the necessary wave coupling between the two wave guides For additional information on microwave diffraction, *see* DIFFRACTION.

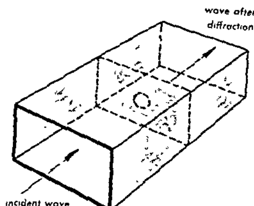


Fig 2. Diffraction of microwaves through a small aperture between two wave guides. Only magnetic lines of force are shown for the field pattern.

Polarization. The polarization of an electromagnetic wave is specified by the direction of the electric and magnetic intensities. For simplicity, consider a plane wave propagating in the z direction. The electric and magnetic intensities are always mutually perpendicular to each other in the xy plane. It is, consequently, necessary to consider the polarization of the electric intensity alone. If the

electric intensity is polarized along one direction, say the x axis, then the wave is said to be plane polarized. If there are components E_x and E_y equal in amplitude but differing in phase by 90°, the wave is circularly polarized (right-handed for E_y lagging E_x and left-handed for E_y leading E_x). Lastly, if the wave is neither plane polarized nor circularly polarized, it must be elliptically polarized. *See* POLARIZATION OF WAVES; POLARIZED LIGHT.

The preceding essentially optical description holds true for microwaves in free space or wherever there is a TEM wave. (For information on TEM waves, *see* MICROWAVE.) Since a hollow wave guide does not support a TEM wave, the description of the polarization of the wave is much more complicated. However, the general notions expressed here are still valid and useful. For instance, the electric intensity of the dominant mode (TE_{01}) in a rectangular wave guide is plane polarized. This situation is also approximately true for the dominant mode (TE_{11}) in a circular wave guide. A circular wave guide is particularly useful in transforming plane-polarized electric intensity into circularly polarized electric intensity by a technique equivalent to the use of a quarter-wave plate in optics. For this purpose, it suffices to use a thin slab of dielectric material, such as mica, to introduce a 90° phase shift for one of the two equal components of the electric intensity.

Faraday effect for microwaves. In optics, the Faraday effect is the rotation of the plane of polarization of a light beam which propagates in a dense transparent medium placed in a magnetic field along the direction of propagation (*see* FARADAY EFFECT). In microwaves, a similar phenomenon has been discovered and has led to many interesting applications.

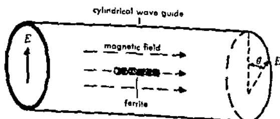


Fig. 3. Rotation of the plane of polarization of a microwave by a ferrite rod in a longitudinal magnetic field. E represents electric intensity.

Consider, for example, a circular wave guide which contains a slender rod of *ferrite* (a magnetic material of very low conductivity), as shown in Fig. 3. Place a steady *magnetic field* along the axial direction. If a wave with vertical electric polarization is incident from the left, then in passing through the ferrite zone its plane of polarization is rotated by an angle θ as indicated in the figure. This Faraday effect can be explained by the action of precessing elementary magnets in the magnetized ferrite upon the phase of the propagating wave. The initial plane-polarized electric intensity is equivalent to two oppositely rotating circularly polarized components. Only one of the components is affected (principally in the phase factor) by the precessing magnets because the latter can only precess in a unique direction corresponding to a given direction of the magnetic field. In the output, the combination of two circularly polarized waves with a relative phase shift is equal to a plane-polarized wave having its polarization rotated by an angle with respect to the initial plane.

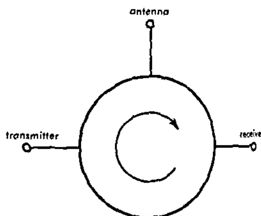


Fig. 4. A "circulator" formed by one or more gyrator elements for nonreciprocal application.

If the output wave is sent back to pass by the ferrite from right to left, the plane of polarization of the backward wave rotates another θ degrees in the same direction as the initial rotation. In other words, the polarization of the new output wave on the left is rotated from the initial plane by 2θ, instead of zero degrees as might be expected by the principle of reciprocity. Thus this system constitutes a nonreciprocal circuit element and is sometimes called a gyrator to signify the gyrating motion of the elementary magnets. *See* GYRATOR; RECIPROCITY, PRINCIPLE OF.

One of the most important applications of a gyrator is found in unidirectional transmission. Suppose a plane-polarized wave is introduced from a rectangular to a circular wave guide containing a gyrator and finally to another rectangular wave guide oriented to suit the polarization of the output wave. If the angle of rotation, which defines the output polarization, is adjusted to 45°, any wave from the output which is fed back to the input will be at right angles with the input intensity and hence will not be accepted by the rectangular wave guide. Another important application of gyrators is known as a circulator, a simple example of which is given in Fig. 4. By means of one or more gyrators, a circulator such as the one depicted will allow a wave from the transmitter to go to the antenna but not to the receiver and will let a wave from the antenna go to the receiver but not to the transmitter. [C.K.J.]

Bibliography: See MICROWAVE.

Microwave spectroscopy

This branch of spectroscopy is concerned with the methods and techniques of observing and the theory for interpreting the absorption and emission of electromagnetic radiation at microwave frequencies by matter—solids, liquids, and gases. There is no universal definition of the frequencies comprising the microwave spectrum, but they are bounded at the low end by radio frequencies and at the high end by infrared frequencies. The frequencies that are used are determined by operating convenience within the approximate range of 10^9 sec^{-1} to 3×10^{11} sec^{-1}, or in wavelength from 30 cm to 1 mm. *See* MICROWAVE; SPECTROSCOPY.

In this frequency region, vacuum tube oscillators of various forms (klystrons, triodes, traveling-wave tubes, magnetrons) are available as sources of monochromatic electromagnetic radiation. The spectrum is observed by passing this radiation through the matter being studied, and the variation in transmission of the matter is noted as the frequency of the radiation is continuously varied. Alternatively, high-intensity electric or magnetic fields, variable in magnitude, are applied to the sample and the variation of transmission of radiation of a fixed frequency with variation of the static field magnitude and direction is noted.

The techniques of microwave spectroscopy are most generally used to observe spectra that can be interpreted by quantum-mechanical theory in terms of fundamental parameters that describe the system, such as interatomic distance, atomic mass, constants expressing the binding potential energy of molecules as a function of interatomic distance, electric and magnetic multipole moments of molecules or nuclei, expectation values of wave functions at the position of these multipole sources, and for semiconductors, for example, the curvature of energy-momentum surface for the electrons—commonly called the effective mass. The methods of obtaining spectra are also used for quantitative detection and identification of particular matter, but the apparatus and techniques are subtle enough to be used primarily for research rather than for routine analysis. The major technological applications of microwave spectroscopy include the ammonia-beam maser and cesium-beam hyperfine resonance cells. These devices are used for time and frequency standards (*see* ATOMIC CLOCK; MASER). Amplifiers using the negative magnetic susceptibility or negative resistance of nonlinear reactances (ferrites or paramagnetic crystals) are also of use as low-noise-figure or low-noise-sensitivity amplifiers, especially in the microwave region. The operating properties of these devices can be explained by using the results of microwave spectroscopic investigations.

Apparatus. Absorption spectroscopy in the microwave region is simplified by the availability of monochromatic sources: triode oscillators, magnetrons, klystrons, traveling-wave oscillators (see Fig. 1). The dispersion elements of optical and infrared spectroscopy (prisms and gratings) are not needed. These sources are tunable in frequency with some difficulty over a small percentage of their operating frequency. The radiation is transmitted through a transmission line (parallel wire, coaxial line, or wave guide) or through a resonant cavity containing the matter to be analyzed. The radiation is detected after it traverses the matter in the absorption cell, and the attenuation caused by the matter is noted.

The signal arising from the absorbed power is generally small compared with that obtained in the infrared region, since the absorption signal strength varies at least as the square of the frequency being observed. The transition probability for radiation-induced transitions is explicitly frequency independent; however, the power absorbed is proportional to the photon energy $h\nu$, where h is Planck's constant and ν is the radiation frequency, and to the difference between induced emission and induced absorption. The latter factor is due only to the difference in the number of systems, for example, atoms, molecules, nuclei, lattice vibrations, in each of the two energy levels between which transitions are being induced. For example, a system in thermal equilibrium with a Boltzmann distribution will have a difference in population in two energy levels separated in energy by $h\nu$ that is proportional to $1 - \exp(-h\nu/kT)$, where k is Boltzmann's constant and T the absolute temperature, or approximately $h\nu/kT$, when $h\nu \ll kT$. To detect this small signal, the absorption is modulated by imposing slowly time-variant electric or magnetic fields on the matter being observed. The signal resulting from the modulation of the radiation by the time-variant absorption is then amplified, detected, and displayed on a milliammeter, oscilloscope, or strip-chart recorder.

Resolution. Since monochromatic sources are available, the line width of the resonant absorption is controlled by physical rather than by instrumental effects. The time τ_1 for the system to come to thermal equilibrium is established, for example, through translational collision in gases, spin-lattice interaction in solids, or spin-spin interaction in liquids, and it determines, through Heisenberg's uncertainty principle, a line width, $\Delta\nu = 1/(\pi\tau_1)$ (see Fig. 2). Doppler broadening in the absorption line is caused by the random motion of the absorbing system in a direction parallel to the propagation of the electromagnetic radiation and results in a line width $\Delta\nu = \nu_0 (\bar{v}/c)$, where c is the ve-
[illegible]

one level to another at a rate sufficient to broaden the absorption with experimental conditions that allow moderate resolution. This unique type of line broadening, saturation broadening, is also observed in the microwave region. The density

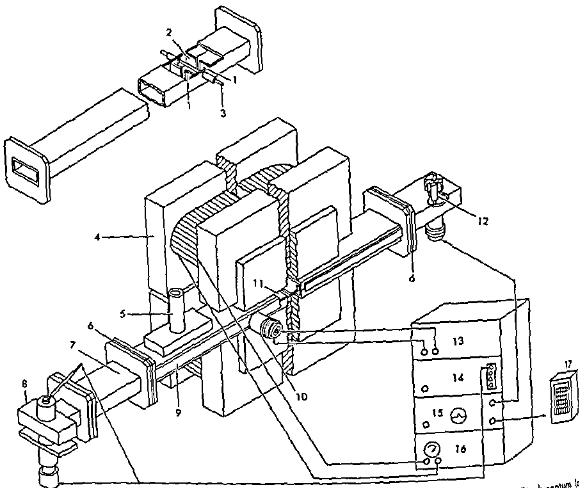


Fig 1. Typical microwave spectroscope using transmission wave guide or transmission cavity: 1, coupling irises; 2, resonant cavity; 3, sample holder; 4, electromagnet; 5, tube to vacuum pump and gas sample handling system; 6, transparent vacuumtight "window"; 7, wave-guide transmission line; 8, klystron; 9, sample-containing wave guide; 10, electrode for introducing electric potential to Stark septum; 11, Stark septum (a flat metal plate for producing electric fields in gas); 12, crystal detector; 13, Stark potential generator; 14, klystron power supply; 15, amplifier with oscilloscope display; 16, electromagnet current supply; 17, strip-chart recorder.

liquids and solids is sufficiently high so that additional broadening arises from internal electric and magnetic fields, or from the inhomogeneity of these internal fields. With molecular beams, however, a resolution $\nu_0/\Delta\nu$ equal to several million can be obtained, and a resolution of 250,000 can be conveniently obtained from resonant absorption in a gas at a pressure of 10^{-2} mm of mercury in cells designed to eliminate all broadening except Doppler width (*see* MOLECULAR BEAMS). Doppler broadening can be reduced by introducing an inert gas with a pressure of a few millimeters of mercury. This reduces the mean translational speed of the absorbing molecules to a diffusion speed; hence, a resolution, $\nu_0/\Delta\nu$, of 10^{9} is obtainable. In liquids and solids, resolution is limited by internal fields to less than 10,000.

The fractional absorption coefficient α, defined as the diminution per centimeter of the intensity of a plane electromagnetic wave, varies from 10^{-1} per cm to a minimum detectable signal of approximately 10^{-11} per cm. It is dependent upon whether the absorbing medium has an electric or magnetic dipole moment, the density of absorbing particles, the number of quantum-mechanical levels comprising the whole system, and the actual frequency. The sensitivity of emission spectrographs is usually, and most conveniently, stated in terms of the minimum temperature variation of a black-body source that is detectable by the apparatus. With spectrometers that have a long response time (the order of minutes), the minimum detectable temperature variation was 0.01K° (early in 1960). Interest in emission spectrographs for radioastronomy and for other uses, however, is bringing about improvement of this figure.

Applications. Microwave spectra of atoms can be used to measure directly and with great precision hyperfine-structure intervals, effective atomic electronic *g*-factors, and nuclear electric and magnetic

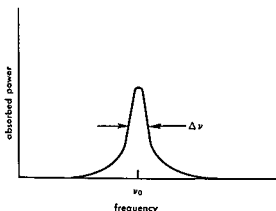


Fig. 2. Absorption curve

moments. The appearance in a microwave spectrometer of its characteristic absorption signal gives evidence of the presence of particular atomic species and can be interpreted quantitatively in terms of atomic density. Such a signal enables the study of atomic energy levels, as well as effects measured by the presence of the atoms and their attenuation with time and with apparatus geometry, such as surface and volume recombination.

The rotational spectra of molecules are dependent, primarily, upon the principal moments of inertia of the molecules, and thus are determined by molecular geometry and atomic masses. Each isotopic form of the molecule will have a resolved spectrum, and interatomic distances of the molecule can be determined from the known atomic masses and the measured isotopic moments of inertia. Since the electron-proton mass ratio is only 1836, the contribution of the electronic moment of inertia to the total molecular moment of inertia must be considered; it is determined empirically from the Zeeman effect, the perturbation of the spectrum by an external magnetic field. The spectrum will be different for varying vibrational excitation, since the average positions of the constituent nuclei vary with this excitation. Although the spectrum is a complex function of the nuclear mass, it has been possible to measure isotopic mass differences to 10^{-2} atomic mass units.

Isotopic abundance and vibrational frequencies can be determined from spectral intensity ratios; the isotopic ratios are given directly by the intensity ratio, and the vibrational intensity ratio is given by the Boltzmann factor, exp $(h\nu/kT)$, where ν is the vibrational frequency. The nuclear electric and magnetic moments interact with the internal molecular field so that nuclear electric quadrupole and magnetic dipole moment fine structure can be observed. Molecular electric and magnetic dipole moments are measured by observing the Zeeman effect and the Stark effect, the latter being the perturbation of the spectrum by an external electric field.

Fine structure arising from inversion in the ammonia molecule and internal torsional motion as in methyl alcohol is sometimes observable, and the interpretation of the resultant spectra gives understanding of molecular binding forces. Studies of absorption-line width and line shape in atomic and molecular spectroscopy have proved useful for understanding intermolecular forces and the part they play in the interchange of energy between separate degrees of freedom—translational, rotational, vibrational, and electronic.

Resolved spectra in liquids and solids arise mainly from paramagnetic or ferromagnetic substances. These systems are weakly coupled to their surrounding liquid or crystal environment and form a distinct energy system with resolvable transitions. Diamagnetic materials, such as semiconductors at low temperatures, have resolvable resonances arising from electrons in conduction states in a static magnetic field These electrons, classically, have circular or helical trajectories, the rotational angular velocity being proportional to the applied magnetic field and inversely proportional to the effective electron mass. Studies of these resonant absorptions have been extremely useful for understanding the symmetry and curvature of the semiconductor energy surfaces, and as a result have contributed greatly to an understanding of the conduction properties and energy band structure of silicon and germanium crystals.

Paramagnetic materials may be formed with ions of the transition elements, with free radicals, or with electrons such as the conduction electrons in alkali metals and the trapped electrons that form crystal *F*-centers. The precision with which the observed spectra of these materials can be described in terms of physical models gives a precise check on the validity and completeness of these models. Although, thus far, microwave spectroscopic data have contributed a great deal of information about spin-spin and spin-lattice relaxation problems, there is still uncertainty about the proper way to describe this interacting system of paramagnetic ions and lattice (*see* MAGNETIC RELAXATION). The theory of the behavior of electrons in metals describes quantitatively only particular aspects of the observed paramagnetic spectra of conduction electrons, for example, the *g*-factor shift. The same comment applies to the field of ferromagnetic, ferrimagnetic, and antiferromagnetic research.

Paramagnetic spectra, besides providing a means for observing interactions in the liquid and solid states, are also finding application as an analytical tool in related fields. The paramagnetic resonance signals from free radicals that appear in life processes provide a means of checking postulated reactions for metabolic processes. The quantitative determination of free-radical density as a measure of their catalytic effect in chemical reactions is another example of the use of microwave paramagnetic spectra in quantitative analysis.

There are applications in electrical engineering, because it has been demonstrated that paramagnetic systems can be prepared so that they have a

negative susceptibility. Interaction of electromagnetic radiation with the paramagnetic system when in this state gives rise to a net emission signal that is coherent with the radiation interacting with it. In this way, amplification can be achieved at the emission frequency. This fact is of particular use in devising microwave amplifiers with high signal-to-noise ratios, or low noise figures. See ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY; MAGNETIC RESONANCE; NUCLEAR MOMENTS; RADIO ASTRONOMY. [M.W.P.S.]

Bibliography: W. Gordy, W. V. Smith, and R. F. Trambarulo, *Microwave Spectroscopy*, 1953; D. J. E. Ingram, *Spectroscopy at Radio and Microwave Frequencies*, 1955; M. W. P. Strandberg, *Microwave Spectroscopy*, 1954; C. H. Townes and A. L. Schawlow, *Microwave Spectroscopy*, 1955.

Microwave transmission lines

Structures used for transmission of electromagnetic energy at microwave frequencies from one point to another. A transmission line may be defined more precisely as a system of material boundaries forming a continuous path from one place to another and capable of directing the transmission of electromagnetic energy along this path. At microwave frequencies, a wide variety of metallic and dielectric structures are used, the choice depending upon the specific application and frequency range. The wavelengths at microwave frequencies are small, ranging from a few millimeters to under 1 meter. It is therefore typical of microwave transmission lines that even when they are physically rather short, their length measured in electrical wavelengths ranges from an appreciable fraction of a wavelength to many wavelengths. Partly for this reason, if microwave transmission lines are not carefully designed, substantial losses of energy by radiation, reflection, and attenuation may be encountered.

Circuit elements used in electrical networks at ordinary radio frequencies, such as coils and capacitors, become so small and inefficient as to be impractical for applications at microwave frequencies. Circuits and networks at microwave frequencies are therefore usually designed with lengths of transmission line serving as circuit elements. These microwave circuit elements can be highly efficient. Using transmission lines, it is possible to construct networks for microwave frequencies equivalent to networks constructed with coils and capacitors that function at lower frequencies.

Microwave structures. The structures in most widespread use as microwave transmission lines are coaxial lines and hollow-pipe wave guides. Other structures, including striplines, open-wire lines, and dielectric rods, are used in special applications. For a detailed discussion of the properties of wave guides, see TRANSMISSION THEORY AND METHODS; WAVE GUIDE.

Coaxial transmission lines. Coaxial lines are widely used at lower frequencies, and are satisfactory for many applications at microwave frequencies. In a coaxial line (Fig. 1) the electromagnetic waves are transmitted through the dielectric medium bounded by two conducting, coaxial cylinders. Because of skin effect, the currents in the conductors are concentrated in the surfaces of the conductors bounding the dielectric medium. To permit transmission of energy in only a single mode of propagation, it is necessary to restrict the mean circumference to less than 1 wavelength. At higher frequencies and shorter wavelengths, the maximum permissible dimensions are small and the losses relatively high. For these reasons, the hollow-pipe wave guide and other types of transmission lines are increasingly preferred at the higher frequencies.

In a coaxial line, the inner conductor may be physically supported by a continuous solid dielectric, as in a flexible cable, or if the dielectric losses are excessively large, they can be reduced by supporting the center conductor by spaced dielectric beads. The center conductor can also be supported by sections of shorted transmission line ¼ wavelength long, called stub supports. Because of their resonant properties, these stubs do not interfere with the propagation of electromagnetic waves along the line.

Hollow-pipe wave guides. Wave guides are extensively used as microwave transmission lines. Electromagnetic waves are transmitted through the interior of hollow metal pipes, and electric currents flow on the inner surfaces. In contrast to coaxial lines, hollow-pipe wave guides are characterized by cutoff frequencies; that is, for a wave guide of given dimensions, the operating frequency must be higher than a critical cutoff frequency for energy to propagate. Wave guides therefore become impractically large at lower frequencies, and the dimensions are reasonably small only in the microwave range.

Although a metal pipe of any cross section will function as a wave guide, a rectangular cross section is most extensively used. Wave guides of circular and ridged cross section are also in widespread use (Fig. 2). The principal advantages of wave guides as compared to coaxial lines are their lower attenuation and structural simplicity.

Microwave striplines. Dielectric rods and tubes can also be used as microwave transmission lines.

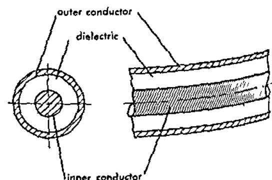


Fig. 1. Coaxial transmission line.

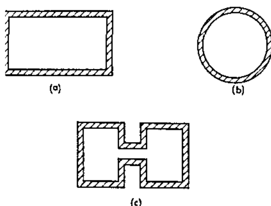


Fig. 2. Typical wave guide cross sections. (a) Rectangular. (b) Circular. (c) Ridged.

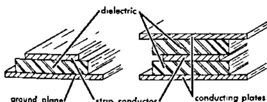


Fig. 3. Microwave striplines.

as can single-conductor lines, which are sometimes coated with dielectric. These structures can have very low attenuation. The electromagnetic fields, however, extend outward from these transmission lines, and serious problems are presented by supports, bends, and shields, which inevitably perturb the fields.

Striplines (Fig. 3), which consist of a metal strip supported above a metal plane or between two metal planes, have somewhat higher attenuation than hollow-pipe wave guides. They are, however, frequently used where structures of unusual compactness and physical simplicity are desired.

Traveling and standing waves. A microwave transmission line performs its primary function of transmitting electromagnetic energy by guiding electromagnetic waves along the line. Normally, the energy is carried along the line in a single mode of transmission (*see* TRANSMISSION LINES). For any transmission line, an infinite number of modes can be excited or established on the line, each characterized by a unique configuration of electric and magnetic fields. The dimensions of the line are usually chosen, however, so that at the operating frequency the line is capable of efficient energy transmission in only a single mode. When the fields of the other modes are excited, they decay rapidly from the point of excitation, and energy is not transmitted any appreciable distance along the line in these modes. If the dimensions of the transmission line are such that energy can be efficiently transmitted in several modes, the structure by which energy is delivered onto the transmission line is usually designed so that only a single one of the propagating modes is excited.

An electromagnetic wave traveling along a transmission line is weakened or attenuated with distance because energy is dissipated in resistive losses in the conductors and dielectric losses in the dielectric media. The magnitude of this attenuation is one of the important characteristics of a transmission line. For an efficient transmission line, the attenuation is low.

Another important characteristic of a transmission line is its phase velocity, computed by multiplying the wavelength of the traveling wave on the line by the operating frequency. Because the wavelength on the transmission line may be greater or less than the wavelength in free space, the phase velocity on the line may be either greater or less than the velocity of light in free space.

A third important characteristic of a transmission line is its characteristic impedance, defined as the input impedance to a uniform line of infinite length. A unique numerical value of impedance cannot be specified for many microwave transmission lines, because no unique definitions of voltage between conductors or total current in conductors exist. However, the ratio of the characteristic impedances of two different transmission lines is often more easily computed, and the knowledge of this ratio is sufficient for many purposes.

The attenuation along a transmission line varies with changes in the operating frequency. For some types of transmission line, the characteristic impedance and phase velocity also change with frequency; for other transmission lines they are nearly independent of frequency.

Waves can travel along a transmission line in either direction. The presence of two waves on a transmission line, traveling in opposite directions, gives rise to standing waves on the line as shown in Fig. 4. If a wave is launched on a transmission line, it continues to travel along the line as long as the line is uniform in cross section. If the wave encounters a discontinuity, or abrupt change in impedance, part of the wave is usually reflected from the discontinuity back toward the source of the wave. A standing-wave pattern then exists between the source and the discontinuity. If the terminal or load impedance of a transmission line is equal to the characteristic impedance of the line, all of the energy in the wave traveling toward the line is absorbed in the load impedance, and the line is said to be matched (*see* STANDING WAVE; STANDING-WAVE DETECTOR). If the load impedance is not equal to the characteristic impedance, part of the traveling wave is reflected. The reflection coefficient is the ratio of the intensities of the reflected and incident waves. When the load impedance is incapable of dissipating energy, as in a short-circuit, an open-circuit, or a pure reactance load, all of the incident wave is reflected. The standing-wave pattern on the line is then of maximum amplitude, and the reflection coefficient is unity.

A discontinuity may exist on a transmission line at some point between the source and the load impedance, causing a partial reflection of the

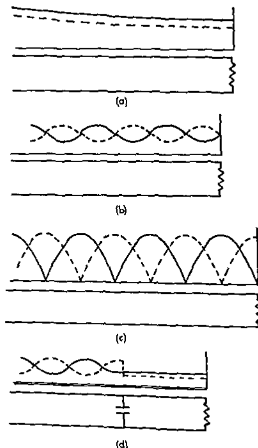


Fig 4 Standing waves on transmission lines. (a) Impedance matched to load impedance, (b) small mismatch, (c) large mismatch, (d) matched to load with discontinuity. Solid lines, voltage; dotted lines, current.

ing wave back toward the source. If the load impedance is matched to the transmission line, there are no standing waves between the load and the discontinuity, but standing waves exist between the discontinuity and the source

Impedance matching. A matched transmission line is one on which only a single wave traveling *from source to load* exists; that is, there is no standing wave. For most applications this is an optimum condition, as it minimizes the undesired losses of energy in the transmission line itself and maximizes the energy delivered to the load impedance. Frequently, however, the load impedance does not match the characteristic impedance of the transmission line, and various matching networks and devices must then be *employed* (*see* IMPEDANCE MATCHING). One such device is a discontinuity with a reflection coefficient equal in magnitude to the reflection coefficient of the mismatched load impedance. If this discontinuity is placed at the proper distance from the load impedance, the reflected waves from the load and from the discontinuity, which are equal in amplitude, will be opposite in phase and will cancel each *other*. The line is then matched between the source and the discontinuity. Another impedance-matching device consists of a section of transmission line ¼ wavelength long connecting the load impedance to the main transmission line. If the characteristic impedance of this quarter-wave section is the geometric mean of the load impedance and the transmission-line impedance, the quarter-wave section matches the load to the transmission line.

One of the most useful impedance-matching devices is the taper. An abrupt transition from one transmission line to another generally produces a mismatch and introduces standing waves. If, however, the dimensions are tapered sufficiently gradually from one line to the other, the reflection is minimized or largely eliminated. Shorter tapers can also be used successfully. A taper that is an integral number of half-wavelengths long introduces the minimum reflection.

Transmission-line structures, such as bends, corners, and junctions, introduce reflections unless they are carefully designed. Transitions from one type of transmission line to another must also be carefully designed if they are to be matched.

Lines as circuit elements. Lengths of transmission line are used as circuit elements in microwave networks. A short length of short-circuited transmission line is equivalent to an inductance. The input impedance of a short-circuited line ¼ wavelength long is very high and is electrically equivalent to a parallel resonant circuit. A short length of open-circuited line presents a capacitive input impedance. If the open-circuited *line is* ¼ wavelength long, the input impedance is very small equivalent to a series resonant circuit.

Graphical solutions. Various graphical charts have been developed as an aid to computation for problems involving transmission lines. Perhaps the most widely used is the Smith chart, developed by P. H. Smith. With this chart, if the impedance at any point on a transmission line is known, relative to its characteristic impedance, the impedance at any other point on the line can be computed Other problems, such as those involving impedance discontinuities and design of matching elements, can also be solved with the aid of these charts *See* MICROWAVE. [T M]

Bibliography: E. L. Ginzton, *Microwave Measurements*, 1957; T. Moreno, *Microwave Transmission Design Data*, 1948; G. L. Ragan, *Microwave Transmission Circuits*, 1948.

Microwave tube

A type of high-vacuum electron tube designed for the generation or amplification of *radio-frequency* power in the range of frequencies from approximately 1000 megacycles/second (Mc) to approximately 300,000 Mc (wavelengths from several centimeters to as low as a millimeter, *see* MICROWAVE). For the lower-frequency portion of this range such tubes may be extensions of the grid controlled triode or tetrode, used extensively at the lower frequencies, with more attention given to the circuit elements within the tube and to the electron paths. However, for most of the range the microwave tubes utilize interaction between electron

beams and traveling-wave or cavity-type microwave circuits. *For tubes designed especially for the microwave range see* Backward-wave tube; Klystron; Magnetron; Traveling-wave tube.

Microwave tubes used as amplifiers may be either for low-level amplification at the input of a microwave receiver, for high-level amplification at the output of a microwave transmitter, or for intermediate levels of amplification between these extremes. Microwave tubes are also used in oscillators to provide the source for microwave-relay, communication, radar, or telemetering systems. For the last two purposes the tubes are normally pulsed, but for the other applications are normally operated continuously. The tubes are also occasionally used in detection and heterodyne mixing, although solid-state devices are more common for such applications.

Characteristics of microwave tubes. Characteristics of importance for all applications of microwave tubes are size, weight, ruggedness, life, reliability, and the number and character of power supply sources required. Other important electrical characteristics depend upon the use. An amplifier used at the input of a microwave receiver will generally handle very low levels of power, so that efficiency is not important. However, the gain or amount of amplification provided and the frequency range or bandwidth over which useful gain is provided are important. For narrow-band amplifiers, such as the klystron, useful gain may exist for frequencies only a fraction of a per cent from the center frequency. For wide-band amplifiers, such as the traveling-wave tube, useful gain may exist for a frequency band up to 50% of the center frequency on each side. In addition to the bandwidth of the gain for fixed adjustments of the tube, the tuning range available by either mechanical or electrical tuning is also of importance. Tubes of the backward-wave class have especially wide ranges of electrical tuning. For an amplifier used with low signal levels, the electrical noise caused by statistical fluctuations in the electron currents is of great importance. If this noise is too high, it will mask the low-level signals which the tube is designed to receive. The quality of an amplifier with respect to this low-noise goal is described by a noise figure F, or by a noise temperature T_n defined by the following

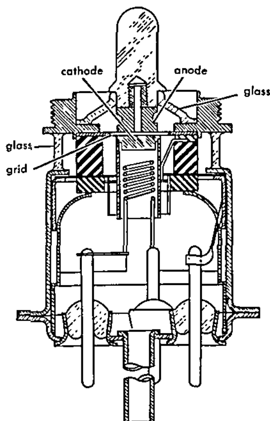
$$F = \frac{\text{Signal-to-noise power ratio at input}}{\text{Signal-to-noise power ratio at output}} \qquad (1)$$

$$T_n = 290(F - 1) \text{ degrees Kelvin} \qquad (2)$$

Noise figures as low as 2.25 (noise temperatures of 360°K) were obtained for microwave tubes in 1958. This is good for most purposes, although not competitive with those available from the maser or the parametric amplifier, either of which may be required in the most sensitive receivers. *See* Noise, electrical.

For the microwave tube used as the generator of power for transmission in a microwave communication, radar, or other similar system, the amount *of power generated, the efficiency of conversion of* the dc power to radio-frequency power, the tuning range and frequency stability, and the capability of modulation are all important characteristics. If the oscillator operates at a low-power level and is followed by a high-power amplifier, as in the master-oscillator-power-amplifier (MOPA) system, the efficiency of the low-power oscillator is not as critical as is the efficiency of the high-power amplifier of such a system. Frequency stability of microwave oscillators is generally better than for corresponding low-frequency oscillators without crystal control, because of the high quality factor Q of the cavity-type circuits used at microwave frequencies (*see* Cavity resonator). For still greater stability, crystal control can be adapted to the microwave frequencies by utilizing higher harmonics obtained from a lower-frequency crystal-controlled oscillator, compared through automatic frequency control (AFC) circuits. Extremes of stability are obtained from the maser-type devices.

Noise arising from the statistical fluctuations in the electron flow is generally not of importance in high-level amplifiers or oscillators, because the noise signals are of low level compared with the desired signal. However, erratic variations of the output caused by microphonics or modulation of the output power by extraneous signals may cause



Vacuum tube 416 A

difficulty. Noise problems in oscillators are generally most important for the local oscillator used in a superheterodyne receiver, because the noise voltages are introduced at a low level in the system and cause interference with the desired signal.

Limitations of conventional tubes. Tubes of conventional construction for the lower radio frequencies (triodes, tetrodes, and so on) will generally not work at microwave frequencies, partly because of the circuit and partly because of transit-time effects in the electron stream within the tube. The circuit limitations come from the interelectrode capacitances, the lead inductances, and the stem capacitances which form low-pass filters and are cut off in the microwave range. Improvement in this respect can be made by proper design of the electrodes, leads, and envelope, but there is an upper limit possible in the grid-controlled types.

The second limitation comes because the time of transit of electrons from one electrode to another is finite (typically of the order of 10^{-8}–10^{-9} sec). This is a small part of a period for radio frequencies of the order of a few megacycles, but is a large part of a period or even several periods for the microwave frequencies. In the latter case, the electron experiences changing phase of its stimuli as it crosses between electrodes, and in turn induces varying phase contributions of currents in the electrodes which cancel rather than add to the desired output. *See* VACUUM TUBE.

The frequency range of tubes of the triode class can be extended by proper design. In the disk-seal construction, the electrode capacitances and lead inductances are reduced to a minimum and the electrodes blend smoothly into the resonant cavity-type circuits which are used with the tube. Transit-time effects are minimized by reducing the interelectrode spacings to the minimum value that can [illegible]

[illegible] diameter spaced 0.001 in. apart, and the resultant grid is placed 0.001 in. in front of the cathode. Even with such extremes of manufacture, the use of this class of tubes is limited to a few thousand megacycles per second.

Types of microwave tubes. In addition to the klystron, magnetron, traveling-wave tube and backward-wave tube, which are discussed in separate articles, there are many interesting principles of microwave amplification which have been demonstrated experimentally and which occasionally find use in limited application.

Double-stream amplification. This utilizes the interaction between two or more electron streams having slightly different velocity. No circuit is required for this interaction except to couple on and off the beams.

Velocity-jump amplification. As above, circuits are required only for coupling, and gain comes by a proper variation with distance of the dc potential.

Resistance-wall and reactance-wall amplifiers. In the former, an electron stream interacts with a purely resistance wall, producing a growing wave as well as the obvious attenuating wave. In the latter, the wall is purely reactive, but not a propagating circuit as in the traveling-wave tube.

Rippled-stream and rippled-wall amplifiers. In these the wall of stream diameter is changed periodically in distance to produce the growing wave.

Interaction between electron streams and plasmas. Interaction between electron streams and plasmas has been used for plasma study for some time, but it also appears that the growing waves from such interaction may also be useful for microwave amplification.

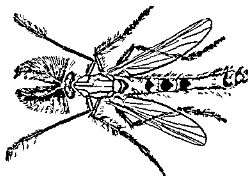
For other microwave devices replacing tubes in certain applications, *see* MASER; PARAMETRIC AMPLIFIER. [J.R.WH]

Bibliography: A. H. W. Beck, *Thermionic Valves*, 1953; C. K. Birdsall and J. R. Whinnery, *J. Appl. Phys.*, 24(3):314–323, 1953; S. Bloom and R. W. Peter, *RCA Review*, 15:95, 1954; A. V. Haeff, The electron wave tube, *Proc. IRE*, 37:4-9, 1949; W. W. Harman, *Fundamentals of Electronic Motion*, 1953; K. Spangenberg, *Vacuum Tubes*, 1948, [illegible]

Midge

Any member of the family Chironomidae of the order Diptera. There are about 200 species in North America. Midges are small, delicate flies, somewhat like miniature mosquitoes in appearance. Most of them have aquatic larvae, red in color, called bloodworms. A few species have larvae which develop in decaying vegetable matter, or under tree bark. The adults frequently occur in large swarms, hovering in a characteristic dancing mass, and making a very audible humming with their wings. After large emergences of midges, they often accumulate in drifts in windows and doorways.

Bloodworms are scavengers and are among the most abundant of bottom organisms in most bodies of fresh water. They are highly important food or



Midge, *Chironomus* sp.; length to ½ in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

ganisms for fishes. Bloodworms are unusual among complex animals in that they are able to survive in lake bottoms where there is no measurable oxygen for weeks at a time. Bloodworms swim by a wriggling movement similar to that of larval mosquitoes. There are several generations a year in most species. Adults of most midges are harmless, but punkies or no-see-ems are blood-sucking midges that sometimes become pests of consequence. See DIPTERA; FLY. [J.D.B.]

Midnight sun

A phenomenon observed in the polar zones of the earth during the summer solstice, when the sun remains visible above the horizon at midnight and reaches its minimum altitude without setting.

The midnight sun is a consequence of the tilt of the rotation axis of the earth, by which the earth presents in turn each pole to the sun for 6 months. The length of the period of uninterrupted daylight decreases as one goes away from the poles and, in principle, would vanish at the Arctic and Antarctic circles (latitude $\pm 66°33'$). However, because of atmospheric refraction of $34'$, the midnight sun can be seen for a few days around the summer solstice from all points beyond $\pm 66°$ latitude. See EARTH (ORBITAL MOTION). [C.D.V.]

Migmatite

Rocks originally defined as of hybrid character due to intimate mixing of older rocks (schist and gneiss) with granitic magma. Now most plutonic rocks of mixed appearance, regardless of how the granitic phase formed, are called migmatites. Commonly they appear as veined gneisses.

Several modes of origin have been proposed. (1) Granitic magma may be intercalated between thin layers of schist (lit-par-lit injection) to form a banded rock called injection gneiss. (2) The granitic magma may form in place by selective melting of the rock components. (3) The granitic layers may develop by metamorphic differentiation (redistribution of minerals in solid rock by recrystallization). (4) The granitic layers may represent selectively replaced or metasomatized portions of the rock.

Veined gneisses include two genetic types: arterites, in which the vein material was injected, and veinites, in which the vein material was secreted from the rock itself. There are many other types of migmatites. Some consist of blocklike masses of various shapes enclosed in granitic rock and resembling fragments of a breccia.

Migmatites are found in zones marginal to intrusive granite and in deep zones of ultrametamorphism. See GRANITIZATION; METAMORPHISM; METASOMATISM. [C.A.CA.]

Migratory behavior

Migrations of animals are regular, usually annual, movements to a different and often a distant environment. Most migrations are associated with annual breeding cycles. Typically, they are between an area suitable for rearing young and another place occupied the balance of the year. Migrations over surprisingly long distances occur in all major groups of adequately studied animals which possess the necessary powers of locomotion. The term migration is also applied to other regular round trips over distances that are long compared to the size and local movements of the animal. For instance, large populations of small planktonic animals of the open ocean perform daily vertical migrations, swimming up and down several hundred feet to remain at approximately the lower limit of penetration of sunlight (see OCEANOGRAPHY).

Annual migrations of 100 miles or more occur in squids, several orders of insects, many fishes, reptiles (sea turtles), the majority of birds, marine mammals (seals and whales), bats, and large terrestrial mammals such as caribou and bison. This is known because of seasonal shifts in populations and by tagging or banding individuals and recovering them after migrations. Many butterflies migrate long distances, especially the monarch (*Danaus plexippus*) which has been marked in North America on a sufficient scale that a few have been recovered hundreds of miles from the place of marking. Most migrating insects are thought to make only a one-way journey, another generation migrating in the reverse direction. Migratory fish may make several round trips in one lifetime, or adults may migrate only once to a breeding area to lay eggs, after which the young migrate in the reverse direction. Many are anadromous like trout, salmon, and smelts, leaving the sea to go to the breeding area upstream in fresh-water streams or lakes. A few, like the eel, are catadromous, leaving the fresh-water rivers and lakes to lay eggs in some part of the ocean. Others migrate long distances through the ocean. So many anadromous fish return to spawn in the same tributaries where they hatched from eggs that the event cannot be explained by random mixing of populations, but it is not known whether eels migrate to the same fresh-water streams as those occupied by their parents.

External and internal factors. Environmental factors giving migratory behavior a survival value include seasonally available food supplies and climates favorable at only one season. It has been suggested that the long days of Arctic summer permit longer hours for feeding the young of birds, and hence, perhaps, shorten the vulnerable nestling period.

Internal factors, important for migratory behavior, include many physiological changes associated with breeding, especially the production of gonadotropic hormones by the anterior pituitary gland, the consequent increase in the size of the gonads and often the deposition of large amounts of fat prior to migration. Particularly in birds, the increasing length of daylight in the spring has a direct, or indirect, effect on the pituitary, causing an increased release of gonadotropic hormones and

consequent growth of ovaries or testes. The enlargement of the gonads, the onset of breeding behavior, and even, in some cases, the spring migration itself can be induced out of season by an arti-

important in bringing on the reproductive state and initiating the annual migration.

Navigational ability. The navigation of migratory animals, the sensory basis of their ability to find their way over long distances, is one of the most important and puzzling aspects of migratory behavior. Migrating fish are thought in some cases to be guided by bottom topography, by swimming or being carried with or against water currents, and in the case of anadromous species by seeking out less saline waters from rivers draining into the sea. Some fish can discriminate the odors in water from one tributary from that of another in the same drainage system, and it has been suggested that this may explain how they locate the particular branch of a river where they were hatched. Some migrating birds may follow coastlines or be guided by wind direction. But birds and many other animals (crustaceans, spiders, insects, fish, and turtles) can orient themselves by the sun They can choose a particular compass direction commonly with a variability of roughly ±30°, but usually cease to do so when the sun is hidden by clouds. Some arthropods can also orient themselves by the polarization pattern of the blue sky. These animals correct for the apparent movement of the sun due to the earth's rotation, utilizing in some way not clearly understood their temperature-compensated "internal clocks" or endogenous activity rhythms. If the animal's time-keeping mechanism is reset by exposure to shifted periods of light and dark, for example, light from noon to 10 P.M and dark for the remaining 14 hours, its directional headings relative to the sun are often shifted accordingly (*see* PERIODICITY IN ORGANISMS).

Most small birds migrate at night, and under certain conditions they will choose the correct direction for their seasonal migration when studied in an experimental cage which excludes any view of local landmarks but allows them to see the night sky or its reproduction by a planetarium. In the latter case, the directional choices are made with reference to the star pattern, and they turn when the optical image of the sky is rotated. The ability to choose the seasonally correct direction of migration may be displayed by birds which have been reared from the egg in laboratory rooms, where they could never have seen the sky or stars. This indicates that recognition of star patterns is based on genetic information. *See* GENETICS; INSTINCTIVE BEHAVIOR. [D.R.CR.]

Bibliography: J. D. Carthy, *Animal Navigation*, 1957; G. L. Clarke, *Elements of Ecology*, 1954; G. V. T. Matthews, *Bird Navigation*, 1955; H. B. Moore, *Marine Ecology*, 1958; C. B. Williams, *Insect Migrations*, 1958.

Milk

The normal secretion of the mammalian mammary gland. In this sense, it serves as food for the young of many species. Man also utilizes milk and milk products as food for adults. Milk for this purpose comes from various animals of which the cow and goat are the chief sources of supply. The cow supplies a large portion of the milk used in Europe and the United States. Commercial sources report total consumption of milk in the United States as 680 lb per capita per year. The name "milk," unqualified, means cow's milk.

This article considers milk, dried milk, skim milk, fermented milk, and concentrated milk. For the discussion of other milk products, *see* BUTTER; CHEESE; ICE CREAM.

The average composition of milk approximates 4% fat, 3.5% protein, 4 8% lactose, 0.7% mineral matter, and 87% water. It also contains vitamins A, C, D, E, riboflavin, thiamin, pantothenic acid, and pyridoxine, as well as certain enzymes, such as phosphatase, amylase, lipase, catalase, peroxidase, and galactase. See ENZYME; VITAMIN.

Fat determination. Where extreme accuracy is important, ether extract methods are used for the determination of the fat in milk. The most common of these methods is the Mojonnier, which enables the operator to combine accuracy and speed. However, because of its simplicity and, for all practical purposes, its accuracy, the Babcock test is commonly used by industry for determining the fat content of milk and cream. Modifications of the test are used for such products as buttermilk, skim milk, ice cream, and condensed milk.

In the Babcock testing procedure, special glassware is used. This includes 18-gram (g) milk test bottles graduated in 0.1% increments from 0 to 8%. For cream a different bottle is used, either 9- or 18-g, with graduations in 0.5% increments from 0 to 50%. Special bottles are also available for testing skim milk, buttermilk, and ice cream mix. For measuring the sample, a 17.6-ml pipette that will deliver 17.5 ml (18 g) is used for testing milk and skim milk. For cream and ice cream testing, the sample must be weighed.

The principle of the Babcock test is to separate the fat in the sample from the nonfat portions of the milk by the addition of concentrated sulfuric acid. A centrifuge is used to facilitate the separation. The fat floated into the graduate portion of the neck of the test bottle is measured by means of calipers. The reading gives the percentage of fat in the sample.

Milk separation and clarification. By running whole milk through a separator bowl rotating at high speed, the cream moves toward the center because of the lower specific gravity of the fat (0.9 as compared to 1.35 for the skim milk), and the skim milk is forced to the outside. To facilitate separation, the bowl contains conical disks mounted on a central spindle. The disks lie close together but far

enough apart to permit the separated cream to move upward and inward to the cream outlet while the skim milk moves upward on the outside of the bowl to the skim milk outlet.

A clarifier bowl functions much the same way except that it contains no dividing wall, hence the only action that takes place is a removal of insoluble dirt, body cells, and a certain amount of protein material. There is only one outlet from a clarifier bowl.

Since World War II, separator bowls have been manufactured that, by a simple adjustment, can be used for clarifying, for separating any test cream desired, or for standardizing the milk to any desired fat content (Fig. 1).

Homogenization. The homogenizer consists essentially of a high-pressure pump which forces the hot dairy liquid (milk, cream, ice cream mix, or evaporated milk) through a small opening, resulting in a reduction in size of the fat globules After milk is homogenized, the fat globules remain in suspension and do not float to the top.

Microscopic examination of milk before homogenization reveals fat globules ranging in diameter from less than 1 to more than 20 microns (μ). After homogenization, the fat globules are more

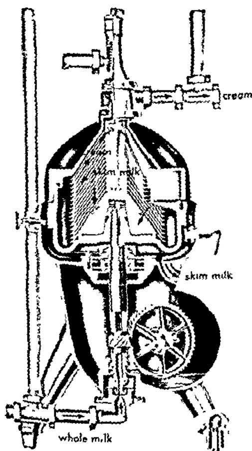


Fig. 1. Cross section of cream separator. (De Laval Separator Co.)

uniform in size and smaller, averaging 1 μ. There are several theories as to the exact cause of the reduction in fat globule size when homogenization takes place. The more commonly accepted theories are (1) the fluid passing at high velocity through the valve, which has a clearance of about 0.0001 in., strikes against the side walls of the valve chamber, causing the fat globules to break apart as a result of the impact; (2) shearing of the melted fat globules occurs when layers of liquid moving at high speed pass through the small opening of the valve and come in contact with layers of slower-moving liquids; (3) the fat globules are flattened and elongated by the acceleration they receive at the entrance to the valve; this is followed by a quick dispersion resulting from the eddy motion, causing a breaking apart of attenuated globules; (4) the cavitation theory explains the reduction in fat globule size as resulting from the shattering that follows the collapse of vapor bubbles formed in regions of high velocity and low pressure. This theory was discredited by the work of A. A. McKillop, W. L. Dunkley, R. L. Brockmeyer, and R. L. Perry in 1955. These investigators believe that the mechanism of homogenization depends upon velocity or velocity gradient and favor the first three theories as satisfying this condition.

Homogenization results in an increase of the amount of protein material adsorbed at the periphery of the fat globules The proteins in milk products are less stable after homogenization. This is particularly true of high-fat products such as coffee cream and evaporated milk. Even whole-milk proteins are destabilized by homogenization, as evidenced by the greater tendency for the curdling of milk in escalloped potatoes, which have a pH lower than that of milk, when the homogenized product is used.

Microscopic examination of homogenized coffee cream and high-fat ice cream mixes reveals a definite clumping of the fat, which causes an increased viscosity. The extent of clumping is dependent in part upon the ratio of fat to nonfat milk solids. Until this ratio is 1:1, or more, the degree of clumping is not significant. Other important factors that favor clumping are high homogenization pressures, use of single rather than double homogenization, homogenizing at temperatures of 57.2°C or lower, and excess of calcium or hydrogen ions. Those conditions which favor clumping also ordinarily lower heat stabilization.

When homogenizing milk in which the lipase has not been inactivated by heat, it is important that pasteurization follow immediately, or a disagreeable rancid flavor will develop. Contamination of homogenized and pasteurized milk with a very small amount of raw milk or cream will also cause hydrolytic rancidity to develop.

The homogenizer has played a very important role in the dairy industry. Its application to market milk has greatly increased the popularity of milk as a beverage and has done much to make

possible the success of the paper milk container. The dealers using only paper containers usually homogenize their entire milk supply. The homogenizer is used to improve the body of coffee cream as well as "half and half," a 10–12% fat product used on cereal and in coffee. Ice cream could not be manufactured as it is today without the use of the homogenizer to prevent fat churning during freezing and to facilitate air incorporation in the freezer. Homogenizing evaporated milk before canning and sterilizing prevents the formation of a thick, leathery fat layer at the top of the can during storage. Homogenization also is an important step in the manufacture of sour cream.

Milk standards. Both chemical and sanitary standards for milk are invoked by Federal, state, and municipal agencies. Certain minimum percentages of fat and nonfat milk solids are required (3.0–3.8% fat and 8.0–8.5% nonfat milk solids). No preservative or harmful chemical, such as formaldehyde or hydrogen peroxide, can be added.

Sanitary standards are primarily requirements pertaining to the health of the cows, the conditions under which the animals are kept, and the methods used in milking and caring for the milk. There are also sanitary standards pertaining to the transportation of the milk to the plant as well as to methods of processing, packaging, and delivery to the store or consumer.

Because of the confusion that often results from the overlapping of state and municipal jurisdiction and the use that is sometimes made of city and state regulations as trade barriers, there is a growing interest in the dairy industry to establish a single standard for milk regardless of where produced or what its use (milk, ice cream, butter, cheese, or evaporated milk), with a single law-enforcing agency.

The U.S. Public Health Service Milk Ordinance and Code contains recommendations widely used as a basis for state laws and regulations. The Milk Ordinance requirements for the various grades of milk follow (footnotes omitted):

"*Grade A raw milk for pasteurization.* Grade A raw milk for pasteurization is raw milk from producer dairies conforming with the . . . items of sanitation [specified in the Milk Ordinance and Code]. The bacterial plate count or the direct microscopic clump count of the milk, as delivered from the farm, shall not exceed 200,000 per milliliter . . . .

"*Grade B raw milk for pasteurization.* Grade B raw milk for pasteurization is raw milk which does not meet the bacterial standard for grade A raw milk for pasteurization, but which conforms with all other requirements. The bacterial plate count or the direct microscopic clump count of the milk, as delivered from the farm, shall not exceed 1,000,000 per milliliter . . . .

"*Grade C raw milk for pasteurization.* Grade C raw milk for pasteurization is raw milk which does not meet the requirements for grade B raw milk for pasteurization.

"*Grade A pasteurized milk.* Grade A pasteurized milk is grade A raw milk for pasteurization which has been pasteurized, cooled, and placed in the final container in a milk plant which conforms with the items of sanitation. . . . In all cases the milk shall show efficient pasteurization as evidenced by satisfactory phosphatase test, and at no time after pasteurization and before delivery shall the milk have a bacterial plate count exceeding 30,000 per milliliter, or a coliform count exceeding 10 per milliliter . . . .

"*Grade B pasteurized milk.* Grade B pasteurized milk is pasteurized milk which does not meet the bacterial-count standard for grade A pasteurized milk, and/or the provision of lip-cover caps . . . , and/or the requirement that grade A raw milk for pasteurization be used, but which conforms with all other requirements for grade A pasteurized milk, and has been made from raw milk for pasteurization of not less than grade B quality, and has a bacterial plate count after pasteurization and before delivery not exceeding 50,000 per milliliter . . . .

"*Grade C pasteurized milk.* Grade C pasteurized milk is pasteurized milk which does not meet the requirements for grade B pasteurized milk."

Certified milk. Certified milk as defined by the American Association of Medical Milk Commissions, Inc., is "the product of dairies operated in accordance with the methods and standards as currently adopted by the American Association of Medical Milk Commissions, Inc., and under direct supervision of local Medical Milk commissions or other agencies recognized or approved by the Association." Frequent inspections are made by representatives of the Commission to check for compliance with Commission requirements. The high quality of certified milk can be seen in its bacterial standards. Certified raw milk must have a total count of not more than 10,000 colonies per ml Pasteurized certified milk must have a total count of not more than 10,000 colonies per ml before pasteurization and not more than 500 per ml after pasteurization and at the time of delivery to the consumer. Certified raw milk must have a total coliform count of not more than 10 colonies per ml and, after pasteurization, not more than one colony per ml.

There is no other food industry as thoroughly or completely regulated as the dairy industry. This has done much to establish consumer confidence in milk and milk products and to promote their use.

Addition of vitamins to milk. Vitamin D is commonly added to homogenized whole milk and skim milk and to practically all evaporated milk, primarily as a preventive of rickets in children. In the early 1920s, defective bone structure among babies and young children was common but it is rarely found now. The Council on Foods and Nutrition of the American Medical Association recommends that vitamin D whole milk shall contain 400 USP units of vitamin D per quart and that evaporated milk shall contain 25 USP units per fluid ounce (equivalent to 400 USP units per quart

when diluted with an equal volume of water to restore a normal concentration of the milk solids. *See* VITAMIN D.

The most common method of adding vitamin D to milk is to use a concentrate prepared by dissolving the crystalline form of vitamin D_3 in milk fat, which is then homogenized in a milk medium, canned, and sterilized.

Modified skim milk may have both vitamins A and D added. Some vitamin and mineral fortified milks are sold which contain added vitamins A, D, B_1, B_2, niacin, iron, and iodine. Most states permit the sale of milk with added vitamins. [P.H.T.]

MILK MICROBIOLOGY

Raw milk. Aseptically drawn milk from healthy cows is not sterile. The interior of the udder is open to invasion by bacteria when the opening of the teat comes in contact with the air and fodder. The bacteria present in the udder are distributed internally by their own growth as well as by physical movement. However, only small numbers, averaging about 1000 per ml, of a few types are normally found in aseptically drawn milk, although much lower and much higher counts are often reported. During the milking procedure, bacteria are present in the largest numbers at the beginning and gradually decrease.

Contamination from external sources. There are many sources from which the milk can be contaminated by microorganisms during milking as well as during the subsequent handling of the milk. The most important are given below.

Stable air. This may contain dust in considerable quantities, especially in a dirty stable or when hay is distributed to the animals before the milking.

Flies and other insects. A fly may carry as many as 1,000,000 bacteria. Such a fly, falling into a liter of milk, will increase the bacterial count of the milk by 1000 per ml, even without bacterial reproduction.

Coat of the animal. Soil, feed, and manure adhere to the cow's coat. During the milking process this material falls from the coat and a portion of it gets into the milk. Dry manure is a source of heavy contamination. Moreover, it contains many microorganisms detrimental to the subsequent processing of milk.

Feed. Hay and silage often contain a great number of spores. Milk is easily contaminated with portions of the feed.

Milk equipment. Equipment, such as pails, cans, coolers, pipelines, bulk tanks, and milking machines, is the most serious source of bacterial contamination. A dirty can may add several hundred thousands of bacteria to each milliliter of milk. It is very important that utensils be made without seams and sharp corners to facilitate cleaning.

Milking personnel. If the milking personnel are not in good health or have infections on their hands, pathogenic bacteria may be added to the milk. Milk may serve as a carrier of human pathogens from one person to another.

Cleaning and disinfection. Thorough cleaning of all utensils is necessary to prevent contamination. The equipment is washed with tepid water to remove residual milk, dismantled, and then brushed with an alkaline detergent containing a wetting agent. Not only the equipment must be cleaned, but also all other objects from which contamination can take place, such as cow, stable, and milker.

After cleaning, disinfection must be carried out when possible. It is accomplished by heat (steam) or by sanitizers. Chemical sanitizers for dairy utensils are fast-acting germicides. The most widely used is hypochlorite. In recent years detergent-sanitizers have been developed by which cleaning and sanitizing can be done in one operation.

Kinds of microorganisms. The saprophytic and pathogenic microorganisms in milk are discussed in this section.

Saprophytes. Saprophytic microorganisms live on dead or decaying organic matter. The important ones found in milk and dairy products are presented by taxonomic family.

1. Lactobacillaceae. Species of this family found in milk convert milk sugar into lactic acid and other by-products. These bacteria are cocci as well as rods, and belong to such genera as *Streptococcus*, *Leuconostoc*, *Lactobacillus*, and others. *See* LACTOBACILLACEAE.

Important species are *S. lactis* (Fig. 2), cause of spontaneous souring of milk and widely used for the making of cheese; *S. thermophilus*, found in fermented milks like yogurt; *Lc. citrovorum*, responsible for the butter aroma; and *Lb. casei*, present in cheese.

2. Enterobacteriaceae. These are gram-negative asporogenous (nonsporeforming) rod-shaped bacteria, commonly occurring in the large intestine of animals. Best known are the species *Escherichia coli* and *Aerobacter aerogenes*. Their presence in milk serves as a sensitive index of fecal contamination. *See* AEROBACTER; ESCHERICHIA.



Fig. 2. *Streptococcus lactis* from soured

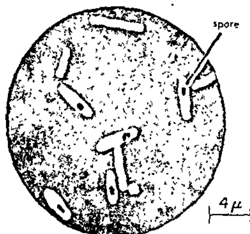


Fig 3. *Clostridium tyrobutyricum.* Anaerobic sporeformer causing blowing of hard cheeses

3. *Pseudomonadaceae.* These are typical bacteria of surface water and are often motile. The genus *Pseudomonas* is well known for causing spoilage, frequently with pronounced biochemical activity, especially on proteins and fats. *See* PSEUDOMONADACEAE.

4. Achromobacteraceae Species are frequently found in soil and water. These bacteria in milk often split fat into glycerol and fatty acids. Some of the volatile lower fatty acids have a sharp odor and impart a rancid taste to milk. Some well-known genera are *Achromobacter* and *Alcaligenes.* The species *Alcaligenes viscolactis* causes ropy milk. *See* ACHROMOBACTERACEAE.

5. Bacillaceae. These are rod-shaped, sporeforming bacteria. The genus *Bacillus* contains aerobic bacilli, like *B. subtilis* and *B cereus.* Because their spores can survive pasteurizing and sometimes even the sterilizing treatment of milk, members of the Bacillaceae are chief causes of spoilage of pasteurized and sterilized milk Some species of the anaerobic genus *Clostridium* attack proteins (*Cl. sporogenes*) and some produce gas (*Cl. butyricum*). They are well known for causing defects in cheese (Fig 3). *See* BACILLACEAE.

6. Yeasts sometimes ferment carbohydrates and produce gas, and sometimes are lipolytic (hydrolyze fats). They occur as contamination in sour milk products and butter and on cheese rinds. *See* YEAST.

7. Molds which actively dissimilate carbohydrates, fats, and commonly proteins often spoil milk and dairy products, but some are useful. For example, *Penicillium roqueforti* is used in making blue-veined cheese.

8. Viruses are submicroscopic agents. In dairy microbiology, interest centers mainly on bacteriophages, which are viruses that parasitize bacterial cells. Bacteriophages attacking lactic acid bacteria can prevent the normal souring which occurs during cheese making. *See* BACTERIOPHAGE; VIRUS.

Pathogens. The milk of diseased animals may contain living germs of pathogenic microbes, and the consumption of such milk by other animals and humans may then cause the disease to be transmitted. Tuberculosis, brucellosis (originally associated only with goats, but later also with cows), Q fever (caused by *Coxiella burnetti* and first reported from Australia), foot-and-mouth disease, and mastitis (a severe disease of the cow's udder) all may be so propagated.

In addition, milk of healthy animals may become contaminated with pathogens of other origin. Milk has been known to transmit in this manner typhoid and paratyphoid fever, septic sore throat, diphtheria, and scarlet fever.

As a rule pathogens do not thrive well in milk but remain alive a fairly long time. They are easily destroyed by pasteurization.

Growth of microorganisms. Milk freshly drawn has bactericidal properties. It contains certain substances, probably body fluids, which destroy microorganisms to a greater or lesser extent. The bactericidal action of milk is weak and is destroyed by heat. Bacterial growth is inhibited for only a short period of time after the milk is drawn. The multiplication of bacteria that have gained entrance to milk also depends on their specific nature, on the numbers initially present, and on the temperature (see table).

Changes in numbers of bacteria (organisms/ml) in milk at various temperatures*

Temp. of holding, °C	Fresh	24 hr	48 hr	72 hr
Clean milk				
4 4	4,295	4,138	4,566	8,1[illegible]
10	4,295	13,961	127,727	5,725,2[illegible]
15.6	4,295	1,587,333	33,011,111	326,500,00[illegible]
Higher count milk				
4.4	136,533	281,646	538,775	749,0[illegible]
10	136,533	1,170,546	13,662,115	25,687,5[illegible]
15 6	136,533	24,673,571	639,884,615	2,407,083,33[illegible]

* Abridged from S H. Ayers, L B. Cook, and P. W Clemmer, USDA Bull. 642, 1918

This table also clearly demonstrates the effect of cooling. It is necessary to hold milk for varying periods on practically all dairy farms. Especially in summertime, cooling milk to below 10°C is necessary to prevent extensive growth. Even this treatment is not always effective, however, because milk may contain psychrophilic bacteria, that is, bacteria that can multiply at low temperatures, and their growth consequently would not be checked.

Quality control. The following microbiological methods are used to control the hygienic quality of the raw milk. The number of living bacteria can be enumerated by the standard plate count. Dilutions of the milk are mixed in an agar medium containing nutrients for the bacteria. The mixture is poured into a sterile glass petri dish and incubated at an appropriate temperature. The number of colonies that grow to visible size are counted. They are

roughly equal to the microorganisms originally present. A direct microscopic count is possible by staining a milk film and counting the stained bacteria in a given area. In the methylene blue reduction test, a correlation is made between the time required to reduce the dye to a colorless form and the population level (number of organisms) of the sample. A similar test is the resazurin reduction test.

There are also tests for specific types of microorganisms, such as for thermoduric bacteria, coliform bacteria, sporeforming bacteria, and also for yeasts and molds.

As bacterial counts are too laborious and expensive, dye reduction tests are generally accepted for grading the hygienic quality of raw milk. However, with low count milk these tests are not adequate and bacterial counts are to be preferred. Specific tests may be of value depending on the destination of the milk.

Liquid milk for consumption. In the dairy plant various processes influence the bacterial population of milk, such as filtration, clarification, separation, homogenization, and pasteurization or sterilization. All the processes, with the exception of pasteurization or sterilization, are of minor importance. Pasteurization or sterilization destroys a large percentage of the organisms in milk, including disease-producing species. There are several processes to pasteurize or sterilize milk, but the one widely used is heating by hot water or steam.

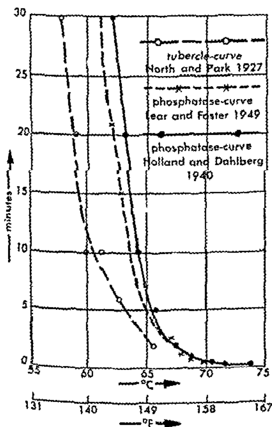


Fig 4 Death curve of tubercle bacteria lies well under phosphatase curve.

Pasteurization. Milk can be processed in bottles, in batches, or on a continuous basis in plate or tube pasteurizers utilizing such time-temperature conditions that pathogens are destroyed with certainty The enzyme phosphatase is present in raw milk, and its inactivation by pasteurization is used as an indicator of the adequacy of pasteurization. Milk that is phosphatase-negative will be free of pathogenic bacteria (Fig. 4).

Although the holding method or vat pasteurization, in which milk is maintained at a temperature of 61 7°C for 30 min, is still being used, continuous pasteurization or high-temperature short-time (HTST) is rapidly gaining popularity. The HTST exposes milk to at least 71.7°C for at least 15 sec. Sporeforming species and so-called thermoduric nonsporeforming species are resistant to this heat treatment. Among the former are *B. licheniformis*, *B. subtilis*, and *B. cereus*; among the latter, *Microbacterium lacticum* and *Streptococcus thermophilus* are the most important. *Microbacterium lacticum* often multiplies in badly cleaned cans.

Spoilage after pasteurization. If no recontamination takes place, for example, with the in-bottle method, spoilage will be mostly due to sweet curdling by sporeformers. *B. cereus* will be the cause if the intensity of pasteurization has been low, because this species is fast-growing. At higher pasteurization temperatures *B subtilis* will be the only remaining sporeformer and consequently the cause of spoilage Contrary to expectation, the keeping quality (KQ) of uncontaminated pasteurized milk decreases if the temperature of pasteurization is increased, because *B. cereus* thrives better in milk which has been heated to a higher temperature. A further increase of the temperature will increase the KQ, because *B cereus* gradually dies out and *B. subtilis*, a slow-growing species, takes over as a spoilage agent

Recontaminated milk, often the result of the bottle-filling process of the HTST method, is spoiled by *Achromobacter* and *Pseudomonas* species when the milk is kept at temperatures below 10°C, and by *S lactis* at room temperature. Because *S. lactis* also thrives better in high-heated milk, the KQ can be decreased again by increasing the pasteurization temperature. An increase of the KQ at still higher temperatures is now impossible, because the presence of *S. lactis* results from reinfection after pasteurizing.

Defects in pasteurized milk are flavor defects caused by *Achromobacter* and *Pseudomonas* species, slimy caused by *Alcaligenes viscolactis*, and bitter cream caused by *B. cereus* that produces lecithinase, an enzyme destroying lecithin which stabilizes the fat globules of the milk. As a consequence small fat clumps appear in the cream layer.

Quality control of pasteurized milk is best effected by keeping-quality tests and subsequent plat-

ing or titrating. The coliform test is also a useful indicator of recontamination.

Sterilization. Sterilization of milk by heat is gaining ground, especially in Europe. Sterilization by radiation has been investigated in the United States, but has not proved practical because of side effects, such as flavor change. While the KQ of pasteurized milk is limited, that of the sterilized milk is meant to be unlimited.

Originally milk was sterilized after bottling. The heat treatment had to be severe and as a result, the color turned brown and the flavor very bad. Further studies indicated that the process can be carried out in two steps. In the first one, a flow sterilization for a short time (1–20 sec) at a high temperature (126.7–148.9°C), practically all microorganisms are killed, even resistant spores. The milk becomes even whiter than before and the taste is not adversely affected. It is possible to reduce the spore concentration from 100,000,000,000 to 1000 per liter (Fig. 5).

The heat treatment is generally carried out in plate or tube sterilizers. However, in some apparatuses the milk is brought directly into contact with the steam (uperization). After this treatment, the sterile milk has to be bottled. At present only tins can be filled aseptically. Even at a filling temperature as high as 80°C it is impossible to fill bottles or paper containers without recontamination; the containers must therefore be resterilized. The degree of contamination depends on cleanliness of the filling machine and the bottles.

During the second step the bottled milk is sterilized again but at a relatively low temperature (for example 115.6°C for a rather long time—15 min). The milk remains practically white and the flavor is not unduly changed.

The sterilizing effect of both sterilizing steps can be controlled by inoculating the milk with known quantities of the highly resistant spores of *B. subtilis* and determining the number of spores before and after treatment.

Spoilage of sterilized milk may occur if the heating in the first step has not been adequate or if the bottles have not been satisfactorily disinfected. *B. subtilis* is the causative organism if the sterilization in the first step has not been effective, and the spoiled milk tastes bitter. *B. circulans* is usually the cause of spoilage if the bottles have not been properly disinfected; a carbolic taint results (Fig. 6). When high filling temperatures are used the milk may be contaminated in the filler with spores of *B. stearothermophilus* and *B. coagulans*. These spores are often not destroyed during the heating in the second step. As a rule they cause spoilage only in warm countries.

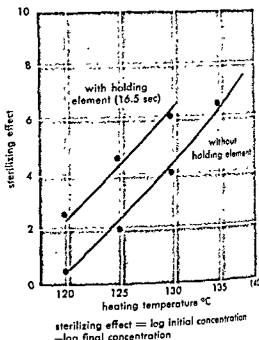


Fig. 5. The sporicidal effect of treatments in an ultra-high-temperature short-time (UHTST) sterilizer.

FERMENTED-MILK PRODUCTS

General procedure. The types of fermented-milk products vary as to the kind of milk and the organisms used in the preparation. Although the origin of fermented milk is ancient only few of the products are produced commercially now. The most important ones are treated here. The same general procedure is used in making all commercial fermented milks. Concentrated, whole, part skim, or skim milk is heated, cooled, and then inoculated with a starter culture and incubated until the desired flavor and body are attained.

Buttermilk. True cultured buttermilk is obtained from the churning of cultured cream in the butter-making process as is done in many European countries. It is also made from sweet cream buttermilk or more commonly from skim, part skim (1.5–2.0% fat), or whole milk by inoculating it with butter starter cultures, incubating at 21°C, and cooling the product at the desired acidity. The fermentation is essentially the same as that preceding the souring of cream (*see* BUTTER). The desired flavor is caused by volatile acids and diacetyl originating from citric acid. To improve flavor, citric acid or common salt is sometimes added to the milk. To give it a likeness to true buttermilk, butter granules are occasionally added. The careful propagation of the starter as a mother culture and bulk culture is highly important in order to retain the original properties of the starter and to ensure the uniformity of the product. The common defects are wheying off, lack of typical flavor, and too high acidity. A bacteriophage can inhibit or prevent acid production in buttermilk just as it can in other fermented products (*see* BACTERIOPHAGE; CHEESE).

Cultured sour cream. This milk product usually contains about 20% fat and is prepared in a manner similar to that for cultured buttermilk.

Acidophilus milk. Acidophilus milk was formerly widely used in the United States and Germany. It is now produced on a limited scale. It was claimed that the growth of *Lb. acidophilus*, under conditions

prevailing in the intestinal tract, will replace undesirable putrefactive fermentations with a beneficial lactic acid fermentation. If acidophilus therapy is to be of any value, large numbers of viable cells must be consumed daily along with a suitable carbohydrate such as lactose or dextrin. The organism grows relatively slowly, and it is easily suppressed by the growth of other microorganisms. It is necessary therefore to sterilize the milk before inoculation. The sterilized milk is inoculated with 1–5% of the culture and incubated 18–24 hours at 37°C.

Yogurt. This is one of the oldest known fermented milks, especially in those countries bordering the eastern Mediterranean coast. It is a popular food in Western European countries and its commercial production is highest of all fermented milks. There are three types of yogurt: Bulgarian yogurt, made from milk concentrated to two-thirds of its volume or with 4–5% of dry skim milk added; and milk yogurt and fluid milk yogurt, both made from whole or standardized milk. Bulgarian yogurt and milk yogurt are custardlike; the fluid milk yogurt is made fluid by stirring. For the custardlike yogurts, the milk is heated for about 5 min at 90–95°C, homogenized, inoculated with 2½% of mother culture, bottled and incubated for about 2 hours. After this, rapid cooling below 10°C is of utmost importance to prevent excess acidity. The mother culture is made of two kinds of bacteria, *Lb. bulgaricus* and *S. thermophilus* (Fig. 7); growth is symbiotic because the rods produce stimulatory substances (amino acids) for the growth of the cocci. In a mixture, the stimulated streptococci grow so rapidly that the clotting of the milk is attained much earlier than with separate strains. In ripe yogurt the proportion between rods and cocci is about 1:1.

The balance between the species is influenced by many factors—the acidity of the culture, the incubation temperature, and time. Antibiotics and bacteriophages to which *S. thermophilus* is very sensitive will inhibit the culture. Consistency of the product and especially the separation of the whey are highly influenced by the culture itself, by the heat treatment, and by the homogenization of the milk. Besides flavor defects (bitter, flat, too acid, unclean), curd firmness defects are frequently encountered, especially with milk yogurt (Fig 8) These defects do not occur in fluid milk yogurt introduced in the Netherlands in the late 1950s. A pronounced slime-forming mixture of yogurt bacteria is used for the manufacture of this type of yogurt. Incubation takes place in tanks for 6–10 hours at 32–35°C. After stirring, the viscous product is bottled. The milk treatment prior to inoculation is the same as for other types of yogurt.

Kefir. Some fermented milk products contain carbon dioxide and alcohol in addition to lactic acid. The best known of these beverages is kefir. The fermentation is started by the use of kefir grains, convoluted masses resembling popcorn. Kefir grains consist of *Lactobacillus casei*, lactic acid streptococci, *Saccharomyces kefir* (a lactose-fermenting yeast), and other microorganisms.

Pasteurized milk is incubated with the kefir grains at 15–20°C for about 8 hours. The grains are removed by sieving and the milk is kept in closed bottles for 24 hours or longer to obtain the desired flavor. The milk becomes effervescent owing to the carbon dioxide produced by the yeast, which is also responsible for the production of the ethyl alcohol. [J.W.P.]

SKIM, CONCENTRATED, AND DRIED MILK PRODUCTS

Skim milk. This is a by-product of the preparation of cream for table use and for the manufacture of ice cream, cultured cream, and butter. Whole milk is centrifuged to remove the milk fat. Skim milk contains about 8.9% milk solids, including 0.1% fat. Much of the skim milk finds an outlet in

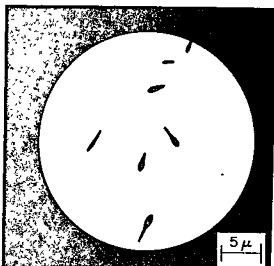


Fig. 6. *Bacillus circulans*, aerobic sporeformer, causing carbolic taint in sterilized milk bottled in improperly disinfected bottles.

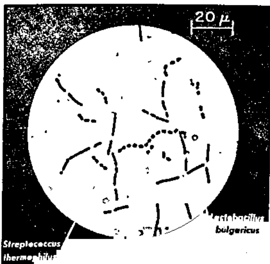


Fig. 7. Photomicrograph of a yogurt culture.

... chocolate drinks, cottage ... s on the ... he consumers' interest in low-fat diets have been helpful in attracting interest to skim milk.

Concentrated milk products. All concentrated milk products are first processed in a vacuum pan to remove a certain amount of water. The milk is preheated to a temperature high enough for proper pasteurization, usually a minimum of 65.6°C but this varies with the product. The milk then passes into a vacuum pan (evaporator), where it boils at a reduced pressure (24–27 in. of vacuum) at a temperature of 48.9–62.8°C. Condensing can be done in single or double effect units. Double effect pans are well suited for continuous operation (Fig. 9). Heat in the pan is supplied by the incoming milk and by steam coils or chests. In a double effect pan, the hot vapors from the first stage supply a portion of the heat needed for the second, resulting in an economy of operation.

A low-temperature evaporator, developed primarily for making concentrated orange juice, is now being used for condensing milk when it is desired to hold the temperature below 32.2°C. With this system the high side of an ammonia refrigeration system is used to supply heat and the low side of the system for condensing the vapor.

Concentrated skim milk. This is used primarily as a source of nonfat milk solids in ice cream. It is perishable and must be kept refrigerated. The percentage of total solids varies from 28 to 32%.

Sweetened condensed skim milk. Sweetened condensed skim milk is used in baking and in ice cream formulas. The milk solids average about 28% and the sugar about 42%. The product is marketed in 55-gallon barrels and 10-gallon milk cans. The added sugar serves as a preservative.

Sweetened condensed whole milk. This concentrated milk is used for candy-making and for household purposes. The composition approximates 8.5% fat, 20% nonfat milk solids, and 44% sugar. The product is packaged in 55-gallon barrels or 10-gallon milk cans for food processing plants and in small tin containers for household use.

The high preheating temperature of 93.3°C and the high sugar content of sweetened condensed skim milk and whole milk result in a product of reasonably good keeping quality even when unrefrigerated. Long storage at room temperature or above causes the condensed milk to thicken and turn brown as the result of a reaction between the milk proteins and sugar. Yeast fermentation and mold ...

3 to 1 concentrate has about the same shelf life as pasteurized milk. It is used undiluted in coffee, on cereal, and as a beverage when diluted with water.

Evaporated milk. This product is a 2 to 1 concentrate of whole milk sterilized in the can at 115.6°C. It is used extensively for industrial and household purposes. The high sterilizing temperature necessary in making the product turns the milk a light brown color and imparts a cooked flavor, which limits its use as a beverage. Its excellent ...

... to remain an important outlet for milk solids.

Filled milk. Filled milk is manufactured for household use in the states where it is permitted. It is made by condensing a mixture of vegetable oil, skim milk, and emulsifying agents until a composition of 6% fat and 26% total solids is obtained. The rest of the procedure is essentially that used for making evaporated milk. It is illegal to ship filled milk across state borders.

Spoilage of evaporated and filled milk may be caused by the survival of a heat-resisting, spore-forming organism, but this occurs only under unusual conditions, such as faulty processing or improper sealing.

Concentrated whey and buttermilk. These byproducts of cheese- and butter-making, respectively, may be condensed in a vacuum pan to a semisolid state and used as such for poultry and hog feed ... in lactose, can be ... ncentrated ... the vacuum pan and used as stock feed. The use of

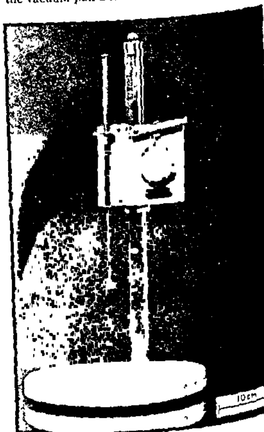


Fig. 8. Consistometer for yogurt.

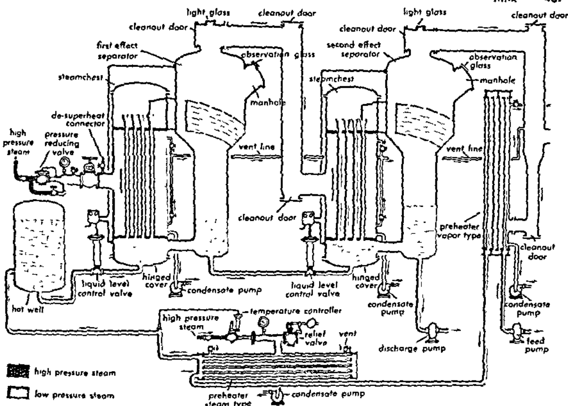


Fig 9. Double effect milk evaporator. (Blaw-Knox)

Lactobacillus bulgaricus culture in the skim milk is helpful in developing the acidity needed to preserve the product.

Dried milk and milk products. Dried whole milk, cream, skim milk, ice cream mix, and modified milk products are successfully manufactured commercially. Because of their low moisture content of 2.5-5.0%, these items are less perishable, can be transported more economically, and have a wider range of usage than the fluid products from which they are made.

A considerable portion of the dried skim milk, whey, and buttermilk—particularly that sold as animal feed—is manufactured on roller dryers. By this process the concentrated liquid is applied in a thin film upon the smooth surface of a continuously rotating steam heated drum. The dried film of milk solids is removed by a sharp, stationary metal blade located opposite the point where the milk is applied. The "sheet" of dried milk is caught in a trough where a spiral rotating mechanism breaks the film into particles and moves them to fillers. Because of its low initial cost, limited space needed, and simplicity and economy of operation, the drum dryer met with early popularity. However, the poor solubility of the finished product, its cooked flavor, and the common presence of burnt particles limit the usefulness of the drum dryer. By operating the roller in an enclosed compartment in which a partial vacuum can be maintained, a superior product results. This method has been used for manufacturing powdered whole milk and malted milk, as well as nonfat dry milk solids.

The most commonly used method for drying milk is spray drying (Fig. 10). This process is well suited for manufacturing products in which low solubility and a minimum of cooked flavor are important. Several commercial spray drying methods have been developed that vary in one or more features but the basic principles are the same. The heated, concentrated milk is forced by a pressure pump through a spray nozzle into an open drying chamber in which hot air, heated to 148.9°C, is moving rapidly. As the hot dry air mixes with the fine mist of heated milk, the moisture vaporizes and the heavier solids drop to the bottom of the dryer. The moisture-laden air leaves the drying chamber through dust collectors. A variant of this procedure is the use of a secondary collector to separate the powder remaining in the air stream from air. The powder then passes through a redryer where the moisture content is further reduced. Since the temperature in the drying chamber is approximately 79.4°C, most systems operate on a continuous basis, the freshly formed powder being removed by a conveyor to be sifted and packaged.

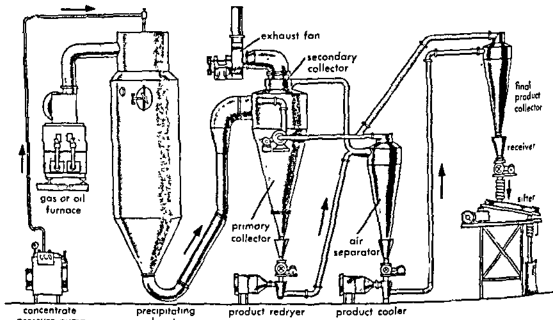


Fig. 10. Lo-Temp evaporator and spray dryer. (*Mojonnier Bros. Co.*)

Instant dried skim milk. Instant dried skim milk can be readily dispersed in cold water. Previously dried milk is put into the loading hopper. The dried skim milk enters the agglomerating section, where it is wetted sufficiently by moist air to cause powder particles to cluster. The clusters, ¼–½ in. in diameter, are separated from the air stream in the collector at the end of the agglomerating tube. The clusters are picked up at the bottom of the agglomerator collector by a high-velocity stream of hot air (132.2–148.9°C). The clusters are dried to the desired moisture content (218–3.6%) in the vertical tube and collector. The redried clusters go from the dry collector to the shaker cooler, a nylon-covered oscillating table measuring 2 ft by 15 ft through which cooling air is forced. The temperature drop

Fig. 11. Equipment for making instant powder. (*Cherry-Burrell Corp.*)

ranges from approximately 73.9 to 43.3°C. The clustered powder passes directly to the sizer, which produces particles of uniform size (Fig. 11). [F.H.T.]

Microbiology. Although concentrated and dry milk products undergo severe heating in the various processes, low-grade milk cannot be improved by these procedures. Good products can only be obtained by starting with a milk of excellent bacteriological quality.

Spoilage. The spoilage of condensed and concentrated milk is essentially the same as that of pasteurized milk. Spoilage of evaporated milk is due to bacteria spores not killed during the heat treatment, bacteria introduced as contaminants during subsequent operations, or bacteria which have gained entrance afterwards because of leaky cans. Curd formation may be caused by sporeformers such as *Bac. coagulans* and *Bac. stearothermophilus.* Other defects are bulged cans caused by gas-producing organisms (*Clostridium* types) or bitterness due to *Bac. subtilis* species. Many of the defects reported in the literature do not occur any more because of the improvements in processing. The reliability of the operations is usually conveniently checked by storing a number of cans at 37°C for several days, after which the incidence of spoilage is determined.

Sweetened condensed milk. This milk product is preserved by the addition of sugar instead of by heat sterilization. After condensing, the sugar concentration will be as high as 65%, which suppresses any microbiological activity due to the high osmotic pressure. Moreover, the sugar concentration is not always as high as 65% and is apt to be much lower in the bulk product, which must therefore be used within a short time. The product is not sterilized and contains a few hundred to over 100,000 microorganisms per gram. Defects of microbial origin are swelling of the cans, caused by yeasts; buttons on the surface, caused by molds if there is too much air in the head space; thickening caused by micrococci and yeasts; and rancidity due to lipolytic microbial enzymes.

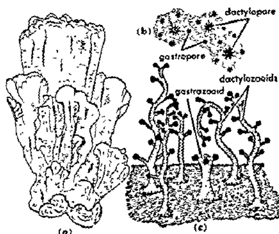
Dry milk products. The microbiology of milk powder is relatively simple because at the low moisture level of this product no microbial metabolism is possible. Microorganisms slowly die, and thus the plate count will drop during storage of the powder.

To obtain the highest plate count the powder must be dissolved at 50°C and the plate incubated at 30°C. A variety of species normally will be found. Occasionally contamination of the storage or feed tank (between condenser and drier) may seed the products with large numbers of thermoduric and thermophilic bacteria, among them staphylococci, which produce toxins. [J.W.P.]

Bibliography: P. R. Elliker, *Practical Dairy Bacteriology*, 1949; E. M. Foster, F. E. Nelson, M. L. Speck, R. N. Doetsch, and J. C. Olson, *Dairy Microbiology*, 1957; O. F. Hunziker, *Condensed Milk and Milk Powder*, 7th ed., 1949; J. W. Pette and H. Lolkema, Yoghurt, *Neth. Milk Dairy J.*, 4:197, 209, 261, 1950, 5:14, 27, 1951; G. M. Trout, *Homogenized Milk*, 1950; USDA, *Standards for the Composition of Milk Products*, Agriculture Handbook 51, rev. 1956; U.S. Dept. of Health, Education, and Welfare, *Milk Ordinance and Code*, Public Health Serv Publ 229, 1953; K. G. Weckel, Theory and practice of vitamin D fortification of milk, *Quart. Rev. Pediat.*, 8:224, 1953; World Health Organization, *Joint FAO/WHO Expert Committee on Milk Hygiene*, Tech. Rept. Ser. 124, 1957.

Milleporina

An order of the class Hydrozoa of the phylum Coelenterata These are the "stinging corals" of shallow tropical seas Their structure is similar to that of hydroids except for the addition of a calcareous exoskeleton. Because of this skeleton, they resemble true corals which belong, however, to a different class of coelenterates, the Anthozoa. The skeleton is covered by a thin layer of tissue which is penetrated by interconnecting tubes and perforated by tiny holes through which the bodies of the "coral animals," or polyps, are extended.



Milleporina. (a) Piece of dry *Millepora*, showing typical flabellate shape. (b) Same, magnified, showing pores. (c) Polyps of *Millepora*. (From L. H. Hyman, *The Invertebrates*, vol. 1, McGraw-Hill, 1940)

The polyps are of two types: nutritive gastrozooids with tentacles and mouth; and protective polyps which are long and armed with stinging cells, but which have no mouth. Their sting is very painful to man, in contrast to the relatively harmless sting of most true corals. Millepores produce medusae, or jellyfish, in which sex cells develop.

Some authorities combine the Milleporina with the Stylasterina in a single order, the Hydrocorallina. *See* HYDROZOA. [S.CR.]

Millerite

A mineral having composition NiS and crystallizing in the hexagonal system. Millerite usually occurs in hairlike tufts and radiating groups of slender capillary crystals. There is rhombohedral c

but it is difficult to observe on the hairlike crystals. The hardness is 3–3.5 (Mohs scale) and the specific gravity is 5.5 The luster is metallic and the color pale brass yellow. Millerite forms at low temperatures, often in cavities and as an alteration of other nickel minerals. It is found in many localities in Europe, notably in Germany and Czechoslovakia. In the United States it is found with pyrrhotite at the Gap Mine, Lancaster County, Pennsylvania; with hematite at Antwerp, New York; and in geodes in limestone at Keokuk, Iowa. *See* NICKEL; PYRRHOTITE [C.S.HU.]

Millet

A name applied to three different groups of warm-season, quick-growing, annual forage plants. Foxtail millet (*Setaria italica*) is grown for supplemental summer hay, making a crop in 60–90 days. It grows to a height of 2–5 ft, depending on soil fertility and moisture supply. Japanese millet (*Echinochloa crusgalli frumentacea*) is taller and coarser than foxtail millet, but is also used for forage in the short summers of northern United States. Pearl millet (*Pennisetum glaucum*), a much taller and coarser-stemmed species, is grown mostly in the southern states and is prized for both harvested forage and pasture when fertilized well and planted thickly. Forage from all three millets should be harvested prior to blooming and before nutritive content begins to decline. [H.B.S.]

In the United States four general types of disease occur on the millets. These include foliage diseases, head mold, smut, and seedling and root diseases.

Foliage diseases usually occur late in the season and, as a rule, damage is slight. *Helminthosporium* and *Cercospora* leaf spot diseases of pearl millet are examples (Fig. 1). Lesions caused by *Helminthosporium* vary in size from brown flecks to large oval or rectangular spots, in contrast to the small dark brown, gray, or tan-centered spots caused by *Cercospora*. Head molds caused by several fungi may seriously affect seed production, particularly in moist weather (Fig. 2). Individual seeds or entire heads may be covered with wooly mycelial mats which are black, orange, or white, depending on the specific fungi present. Smut diseases, common in other countries, also affect the seed in the heads of foxtail and proso millet in the United States. These smuts may be controlled by seed treatment. Seedling blights and root rots, caused by a number of soil-inhabiting fungi, may reduce stands under unusually moist conditions. *See* PLANT DISEASE. [H.A.BO.]

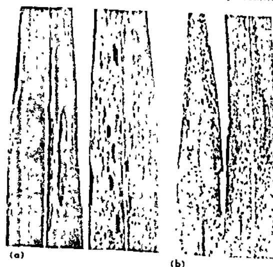


Fig. 1. Leaf diseases of millet. (a) Helminthosporium leaf spot of pearl millet (*Helminthosporium stenospilum* or *H. sacchari*). (b) Cercospora leaf spot of pearl millet (*Cercospora penniseti*). (From E. S. Luttrell et al., *Diseases of pearl millet in Georgia, Plant Disease Reptr.*, 38(7):508, 1954)

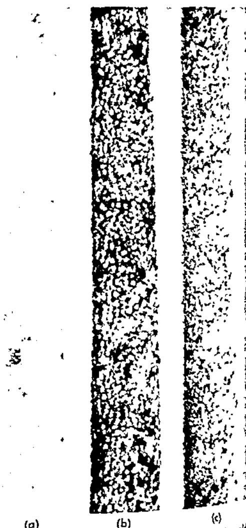


Fig. 2. Head molds of pearl millet (a) Black mold caused by *Helminthosporium* and *Curvularia* (b) Orange mold caused by *Fusarium*. (c) White mold caused by *Oidium*.

Millibar

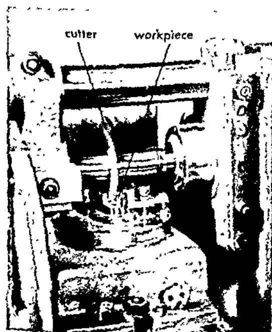
The unit of pressure commonly used in meteorology, a force of 1000 dynes/cm^2 of surface. One millibar is 1/1000 of a bar. The normal atmospheric pressure, a 760-mm column of pure mercury at

0°C in the gravity field at mean sea level and 45° latitude, is equal to 1013.3 millibars. One millibar equals 0.02853 in. or 0.7501 mm of a mercury column. The approximation 1 millibar = ¾ mm Hg gives an error of only 5×10^{-5}. The cgs unit of pressure, the barye, is equal to 1×10^{-3} millibar. Its dimensions are $M\ L^{-1}\ T^{-2}$. *See* ATMOSPHERE; DYNE. [V.E.S.]

Milling machine

A machine for the removal of metal by feeding a workpiece through the periphery of a rotating circular cutter. The multitoothed cutter of a milling machine produces a milled surface as each revolving tooth removes a portion of metal from the passing workpiece.

Milling machines may generally be classed as horizontal or vertical, depending on the power spindle's axis of rotation. A movable, horizontal table holds the workpiece and carries it through the path of the cutter teeth as illustrated. On the



Horizontal milling machine. (Brown and Sharpe Manufacturing Co.)

universal mill, the table may also be rotated horizontally; on machines with a universal head, the cutter may be tilted to the desired angle.

Milling cutters may be made from a single piece of steel or they may be composed of a body with attached cutting elements of a harder material. Various cutters are designed to do specific types of milling operations. *See* MACHINING OPERATIONS. [A.T.]

Millipede

Any member of the subclass *Diplopoda*, class Myriopoda, phylum Arthropoda. There are approximately 7000 species, about 125 of which occur in the United States.



Millipede, Julus sp. (From J. G. Wood, Popular Natural History, Porter and Coates, 1885)

Most millipedes feed on dead and decaying plant material and are of little importance. For the most part they live in damp, dark situations, such as within or under rotting logs and under rocks. Rarely they appear in or under a house in large numbers.

Millipedes are round, wormlike, sharply segmented animals. They are easily distinguished by the presence of two pairs of short legs on each body segment except the first four. Millipedes are mostly dark in color, often gray or bluish gray, sometimes marked with red, yellow, or orange. When disturbed many of them twist into a flattened coil.

Sexes are separate and fertilization is internal. The young have only six somites and three pairs of legs upon hatching, adding others in front of the anal somite at each molt. Like the centipedes, these are often called thousand-legs, or thousand-legged worms. *See* CENTIPEDE; DIPLOPODA. [J.D.B.]

Mineral

A naturally occurring substance, with a characteristic chemical composition, expressed by a chemical formula. Although organic substances such as coal and oil are usually listed under mineral resources, they are not minerals, being complex mixtures without definite chemical formulas. Minerals constitute an extremely important natural resource. Most metals and inorganic chemicals, and many other products essential to civilization, are derived from minerals Both forests and farms are dependent upon soils, which are composed chiefly of minerals.

[illegible]

attached to the surface of some opening. Geodes are cavities in a rock which are lined with a mineral, or in some cases, completely filled. Quartz geodes in limestone are common. They sometimes form well-developed crystals, or they may form fine-grained layers called agate or onyx. The term vug applies to irregular openings containing ore minerals Veins are fissures of varying size which have been filled with one or more minerals. In some cases veins are banded, indicating a sequence of deposition. The term lode is used for a vein or group of veins in a definite area. Gangue minerals are the worthless minerals associated with a valuable mineral or ore. The term ore is usually restricted to minerals from which a metal is obtained. A placer is a concentration of relatively heavy and durable minerals which have been transported and redeposited in a stream bed where the water velocity is lowered. Beach sands

contain a concentration of certain heavy minerals. *See* ORE AND MINERAL DEPOSITS.

Mineral names. Minerals usually have both a chemical name and a mineral name. Thus lead sulfide occurring in nature is called galena, and sodium chloride is called halite. Some old mineral names are of unknown or uncertain origin, while many come from the Latin, as orpiment (*auri pigmentum*) or from the Greek, as chalcocite (*chalcos*, copper or brass).

Most modern names end in "ite." Some have a chemical connotation, as molybdenite, MoS_2, and zincite, ZnO. A crystallographic derivation is illustrated by tetrahedrite and hemimorphite; geographic by labradorite and vesuvianite. Physical properties are the basis for the names magnetite, graphite (to write), rhodonite (rose color), cryolite (ice stone), and azurite. Many minerals have been named after individuals, as scheelite, smithsonite and goethite. Some minerals have variety names, as amethyst, agate, and jasper for quartz. There are also colloquial names, such as fool's gold for pyrite, heavy spar for barite, and tin stone for cassiterite.

Classification of minerals. Minerals are classified first with respect to chemical composition, and then so far as possible by isomorphism, or similarity of crystalline form. In general the cation is of less significance than the anion. Thus the iron minerals pyrrhotite, FeS; pyrite, FeS_2; hematite, Fe_2O_3; magnetite, $FeFe_2O_4$; and siderite, $FeCO_3$, are not grouped together, in spite of the common cation Fe. $FeCO_3$ is classified with the hexagonal carbonate group including calcite, $CaCO_3$, and rhodochrosite, $MnCO_3$. Fe_2O_3 is grouped with the isomorphous corundum, Al_2O_3, and $FeFe_2O_4$ with spinel, $MgAl_2O_4$.

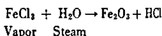
The major groups in mineral classification are as follows:

Native elements	Sulfate type
Sulfide and sulfo minerals	Chromates
Oxides and hydrated oxides	Molybdates
Halogen minerals	Tungstates
Nitrates	Phosphate type
Carbonates	Arsenates
Borates	Vanadates
	Silicates

See BORATE MINERALS; CARBONATE MINERALS; HALOGEN MINERALS; NATIVE ELEMENTS; NITRATE MINERALS; SILICATE MINERALS.

Formation of minerals. Minerals may be formed by four general processes: (1) from a gas by sublimation, (2) from a liquid (aqueous solution), (3) from a liquid (molten rock or magma), and (4) from a solid by metamorphism. In nature these processes may be intimately related.

Sublimation. This process is comparatively rare. In volcanic eruptions minor amounts of certain minerals may form from gases escaping from vents or fumaroles. Examples are salammoniac, NH_4Cl, sulfur, S, and boric acid, H_3BO_3. Scales of hematite may form by the reaction



$$FeCl_3 + H_2O \rightarrow Fe_2O_3 + HCl$$
Vapor   Steam

The only common example of sublimation is unrelated to volcanoes. It is the formation of snow flakes from water vapor.

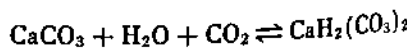
Aqueous solution. Aqueous solutions are the source of important minerals. Dissolved material may be precipitated from solution by several processes: (1) evaporation of the solvent, (2) decrease in temperature and pressure, (3) loss of carbon dioxide, CO_2, or (4) action of organisms.

1. The ocean is a vast reservoir of dissolved materials. During various geologic periods, portions of oceans have been cut off, and subsequent evaporation caused deposition of enormous quantities of the dissolved material. The most abundant of these evaporites are halite (NaCl), and gypsum ($CaSO_4 \cdot 2H_2O$). In some cases evaporation proceeded further, and the more soluble magnesium, Mg, and potassium, K, minerals were also precipitated. In desert regions where occasional rain produces temporary lakes, playa deposits are formed, in which, in addition to halite and gypsum, sodium sulfate and carbonate and various borates occur. *See* EVAPORITE (SALINE).

2. In regions of hot springs and geysers, the hot water under pressure dissolves material from the underlying rocks and brings it to the surface. In Yellowstone Park the geysers bring up siliceous sinter (opal) and the hot springs deposit travertine (calcite). *See* SILICEOUS SINTER; TRAVERTINE.

... limited to

in water, but is slightly soluble if dissolved ... present. The reaction produces $CaH_2(CO_3)_2$ which is soluble.



$$CaCO_3 + H_2O + CO_2 \rightleftharpoons CaH_2(CO_3)_2$$

The many caves in limestones all over the world are formed this way. However, this reaction is reversible. Upon loss of the CO_2, $CaCO_3$ is redeposited in the form of stalactites, stalagmites, and other cave formations. Calcareous tufa may be deposited around springs and streams by this same process. *See* STALACTITES AND STALAGMITES; TUFA.

4. Ocean water does not normally contain enough dissolved $CaCO_3$ or silicon dioxide, SiO_2, to allow much direct precipitation of these substances. But various living organisms are able to extract them from the water, and then secrete them to form the hard parts of their bodies. Corals, crinoids, mollusks, and Foraminifera secrete $CaCO_3$, while diatoms, sponges, and Radiolaria secrete $SiO_2 \cdot xH_2O$. Large deposits of limestone, chalk, and diatomaceous earth have been formed this way.

Crystallization from magma. At moderate depths in the earth's crust, previously existing rocks may be melted. This molten rock (magma) tends to work its way upward. If it reaches the surface there will be volcanic eruptions or lava flows. This

quickly cooled magma forms either glassy or very fine-grained rocks (extrusive) which are of minor mineralogical interest. However, if the magma does not reach the surface, cooling takes place very slowly, allowing time for a complex series of reactions to occur.

In an average magma the major chemical constituents are oxides of silicon, SiO_2; aluminum, Al_2O_3; iron, Fe_2O_3 and FeO; calcium, CaO; magnesium, MgO; sodium, Na_2O; and potassium, K_2O. The most abundant is SiO_2, or silica. In addition, there are substances which would be gases, except for the high pressure, such as water, chlorine, fluorine, carbon dioxide, boron, and sulfur compounds. [illegible]

[illegible] centrated in a phase which is no longer magma, but rather a hot, highly concentrated aqueous solution, called magmatic water. The general order of crystallization from the magma is as follows: (1) basic minerals (low in silica), such as olivine and basic plagioclase, together with nonsilicate minerals, which are oxides and sulfides of metals such as iron, copper, nickel, chromium, platinum, titanium, and diamond; (2) intermediate minerals (medium silica); and (3) acid minerals (high in silica). The igneous rock called granite is the last to crystallize from the magma. See GRANITIZATION; MAGMA.

This progressive change in the kind of minerals crystallizing is called magmatic differentiation. Since the magma tends to move upward, these successive stages may be separated in space.

The residual magmatic liquid is still very hot and concentrated, and under high pressure, but is much less viscous than the magma. It forms pegmatite dikes, frequently as offshoots from the main magmatic mass. It may contain large masses of quartz, big crystals of microcline, and sheets of mica. There may be minerals containing relatively rare elements, such as beryl (beryllium), spodumene (lithium), uraninite (uranium), wolframite (tungsten), and columbite (columbium, or niobium).

The final stage of magmatic differentiation is the hydrothermal stage, with high, medium, and low temperature phases. These occur in veins, usually somewhat removed from the main magmatic mass. Many important minerals occur in these veins, including ores of tin, tungsten, molybdenum, gold, copper, zinc, mercury, and antimony, as well as pyrite, marcasite, quartz, calcite, fluorite, barite, and opal.

Contact metamorphism. Adjacent to the ascending magma the country rock may be profoundly changed by the heat and by chemical reaction with the magma. This is known as local or contact metamorphism. Many new minerals are formed, especially when the magma intrudes into impure limestones. See AUREOLE, CONTACT; METAMORPHISM.

Regional or dynamic metamorphism. Heat and pressure associated with mountain-forming processes, and aided by any water present in the rocks may cause recrystallization of existing minerals or formation of new minerals. Examples of recrystallization are the change of limestone to marble, and a sandstone with siliceous cement to quartzite. Schists may be formed, consisting of minerals with [illegible]

[illegible] formed they are subject to change as chemical processes act upon them. These changes may take place rapidly or they may proceed very slowly.

Weathering of minerals. Weathering may be important in the forming of new minerals, as well as in concentrating them. Deposits of carbonates of copper, zinc, and lead have been formed from primary ores by ground water containing CO_2. Weathering in some tropical regions has produced important deposits of iron, aluminum, and manganese ores. With alternating wet and dry seasons, a special type of leaching occurs. In quite different rocks, practically all constituents may be dissolved except iron and aluminum oxides. The laterite deposits of India are iron ores, while bauxite deposits formed by this process are the chief source of aluminum. Some deposits of manganese oxides have formed this way. Secondary enrichment of certain copper ores has occurred when veins of lean ore have been weathered, with the copper minerals being dissolved and reprecipitated at lower levels. As the surface was gradually lowered by erosion, this process continued until rich deposits were formed. See BAUXITE; LATERITE.

Usually weathering is a destructive force. It is divided into physical and chemical weathering. The first involves changes in temperature, with alternate expansion and contraction; the expansion of water freezing in crevices; the abrading action of rock particles carried by wind, water, and ice. These tend to break down rocks and minerals into finer particles. Chemical weathering involves solution, oxidation, reduction, hydration, dehydration, and carbonation. See WEATHERING PROCESSES.

Pseudomorphs. A special type of weathering may result in pseudomorphs, or false forms. A crystal may be chemically altered, but retain its original shape. Thus a cube of pyrite, FeS_2, may alter to fine-grained goethite, $HFeO_2$, but retain the original cube shape. This is called a pseudomorph of goethite after pyrite, and is an alteration pseudomorph. There may be a complete change in composition, as quartz after fluorite, giving a substitution pseudomorph. In polymorphous substances, crystals of one form may alter to another, but retain the original crystal shape, as rutile, TiO_2, after brookite, TiO_2. These are called paramorphs. In a somewhat different category are crystals of one mineral which have become coated with another. These are called incrustation pseudomorphs.

Metasomatism. Metasomatism is a special type of mineral transformation. For example, in hydrothermal veins where galena has been deposited, galena has been found partly replacing the lime-

stone wall rock, with the original rock texture preserved. *See* METASOMATISM.

Deuteric alteration. This occurs chiefly in pegmatites, where earlier-formed minerals may react with later pegmatite fluids to form new minerals. *See* PEGMATITE.

Mineral groups. Any complete classification of minerals must be based on chemical composition. However, limited groups of minerals may be of interest for various special purposes. Such groups may be based on origin, type of occurrence, certain physical properties, or use.

Primary minerals are those formed from the magma, including the pegmatite and hydrothermal phases. Others are secondary. Rock-forming minerals are those which make up the great bulk of the igneous, sedimentary, and metamorphic rocks. Thus quartz, feldspar, and mica are found in granite and are rock-forming minerals. They are also called essential minerals, in contrast to the occasional tiny crystals of pyrite, zircon, or apatite which might occur in granite, and which are called accessory minerals.

Classes of minerals are those which make up a certain chemical group, as carbonates, sulfates, or oxides.

Isomorphous groups are those whose members are strictly isomorphous, as the garnet group. *See* ISOMORPHISM (CRYSTALLOGRAPHY).

Families of minerals include certain closely related minerals which have close chemical and physical similarities, but are not necessarily isomorphous, as the feldspars or pyroxenes.

Economic minerals are those which are of economic importance, and include both metallic (ore minerals) and nonmetallic minerals, as cryolite and sulfur, and gem minerals. *See* GEM.

Clay minerals have certain common physical and chemical properties. They are fine-grained, plastic when wet, and become hard when dried or fired. They are chiefly hydrous silicates of aluminum. *See* CLAY MINERALS.

Stable minerals are those resistant to both chemical and mechanical weathering, being both insoluble and hard. Heavy minerals are those which collect in placers and beach sands because of their higher specific gravities; or those which can be separated in the laboratory by gravity methods. Detrital minerals are rock fragments which are essentially unaltered. Authigenic minerals are those generated in the place of formation, and allogenic minerals are those which have been transported. *See* AUTHIGENIC MINERALS.

Mineral associations refer to minerals formed by the same process and hence closely associated. Mineral sequence refers to a series of associated minerals formed at successive stages.

A mineral suite is a general term which may apply to a group of associated minerals in one deposit; a representative group from a certain locality; or a group of specimens showing variations, as in color or form, in a single mineral species. *See* MINERALOGY.

[L.S.R.]

Mineral fuel areas

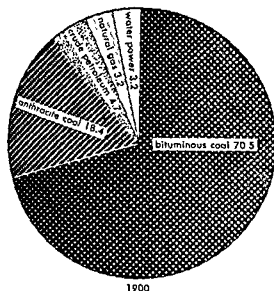
Character and distribution of known resources and resource areas of mineral fuels such as oil, oil shale, gas, and coal, commonly termed fossil fuels. These fossil fuels are organic materials, plant and animal, which in the geologic past accumulated within the muds and sands deposited in the seas to form petroleum and related minerals, or as largely plant deposits in swamps ultimately to form coals and lignites. Thus, through the agency of photosynthesis the sun's heat energy of past ages is preserved in the ground and tapped as major sources of heat and power for present use.

Liquid fuels and reserve areas. Four normal kinds and one synthetic type of liquid fuel resources are considered here. The rising importance of the petroleum resources in the United States is illustrated in Fig. 1.

Liquid petroleum (crude oil). The uneven distribution and widespread occurrence of sedimentary basins, with some likelihood of petroleum reserves, are perhaps best presented on maps such as Fig. 2 (*see* PETROLEUM GEOLOGY). Most other areas are considered unlikely to contain significant amounts of petroleum. The uneven concentration of sizable production and the limited distribution of oil-field areas are patterns emphasized upon maps and tables (see Table 1, Table 2, and Fig. 2).

The amounts shown in Table 1 include crude oil and petroleum gas liquids from shallow seas as well as from the lands. Petroleum gas liquids are liquid hydrocarbons which may be extracted from the gas, that is, separated out of solution in the crude oil when that oil is brought out of the underground reservoir to the lower pressures that exist at the earth's surface. The amount of both gas and associated petroleum gas liquids that occur in crude oil varies in different fields and conditions. The extent of their recovery and use varies with the need for them and the economics of their use in different countries and circumstances. Although there are many shallow-water-covered areas of the world that will yield petroleum, those of greatest present or potential importance are the Gulf of Mexico, California coast, Lake Maracaibo, Gulf of Venezuela, Gulf of Paria, Persian Gulf, Caspian Sea, and the seas off the east coast of Asia and off Indonesia. Areas of continental shelf appear most notably in this list. *See* CONTINENT; CONTINENTAL SHELF AND SLOPE.

Sedimentary basins of the world cover about 22,000,000 mi². This is 36% of the world's 51,250,000 mi² of land surface plus an additional 9,750,000 mi² more of continental shelves and inland waters where the sedimentary basins extend out beyond the shoreline. Not all, however, of the 22,000,000 mi² area is potential oil-bearing basin, as some parts are too shallow or contain unfavorable sedimentary facies for oil occurrence. A reasonable proportion would be 15,000,000 mi² of land areas and 3,000,000 of water areas. This 18,000,000 mi² would be a more realistic total.



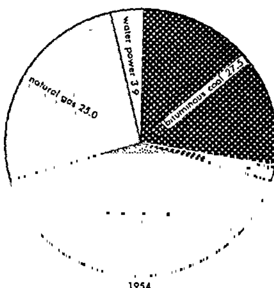


Fig. 1. Graphs illustrating shift in energy sources in the United States from 1900 to 1954 on a percentage basis. On a tonnage basis the coal needed in 1954 was nearly as great as in 1900 because of increased use of power. (From N A. Bengtson and W. Van Royen, Fundamentals of Economic Geography, 4th ed, Prentice-Hall, 1956)

The estimates of reserves and total resources listed in Table 1 include only the oil expected to be produced by conventional primary methods, without the aid of any form of secondary recovery, and under present economics. It should be stated that the estimates of ultimate resources have a lesser basis of accuracy than those of the proved reserves. They do, however, represent what is believed to be an approximate order of magnitude of ultimate petroleum resources in each of the major regions and for the world based on extensive geologic studies

Secondary recovery, using such techniques as repressuring the oil reservoir by means of gas or water injection and other methods of stimulating the flow of fluids from the reservoir rock to the well bore and thence to the surface, greatly increases the yield By such means it is not uncommon to

Table 1. World crude oil and petroleum gas liquids production[a]

Country or area	Cumulative production	% of world	Proved reserves	% of world	Cumulative discoveries	% of world	Ultimate resources[b]	% of world	Daily average production 1959, bbl	% of world
United States	66,597	55.5	36,000	10.7	102,597	22.4	245,000	14.1	7,045,000	36.2
Canada	1,310	1.1	4,600	1.4	5,910	1.3	65,000	3.8	508,000	2 6
Mexico	3,297	2.8	2,500	0.7	5,797	1.3	26,000	1.5	264,500	1.3
Balance N. America	2		3		5		4,000	0.2	500	
Total N. America	71,206	59.4	43,103	12.8	114,309	25.0	340,000	19 6	7,818,000	40.1
South America	15,914	13.3	22,000	6.5	37,914	8.3	140,000	8.1	3,304,000	17.0
Total W. Hemisphere	87,120	72.7	65,103	19.3	152,223	33.3	480,000	27.7	11,122,000	57.1
Europe[c]	793	0.7	1,500	0.4	2,293	0.5	15,000	0.9	250,000	1.3
Middle East[d]	14,631	12.2	220,000	65.2	234,631	51.3	700,000	40 3	4,655,000	23 9
Africa[e]	38		9,000	2.7	9,038	2.0	85,000	4.9	50,000	0.3
Far East[f]	3,490	2.9	11,500	3.4	14,990	3.3	70,000	4.0	487,000	2.5
Total E. Hemisphere	18,952	15.8	242,000	71.7	260,952	57.1	870,000	50.1	5,442,000	28.0
Total Free World	106,072	88.5	307,103	91.0	413,175	90.3	1,350,000	77.8	16,561,000	85.1
Iron Curtain countries	13,887	11.5	30,247	9 0	44,134	9.7	385,000	22.2	2,896,000	14.9
Total world	119,959	100.0	337,350	100.0	457,309	100.0	1,735,000	100.0	19,460,000	100 0

[a] Includes reserves, discoveries, and total resources as of January 1, 1960 in 10^6 barrels (except for daily production) for both land and water areas. [b] Cumulative discoveries to date plus estimated future discoveries. [c] Excluding U.S.S.R. and other Iron Curtain countries. [d] Including Egypt. [e] Excluding Egypt. [f] Including India, Pakistan, Burma, Japan, Indonesia, and other countries of Oceania but not Red China and Siberia.

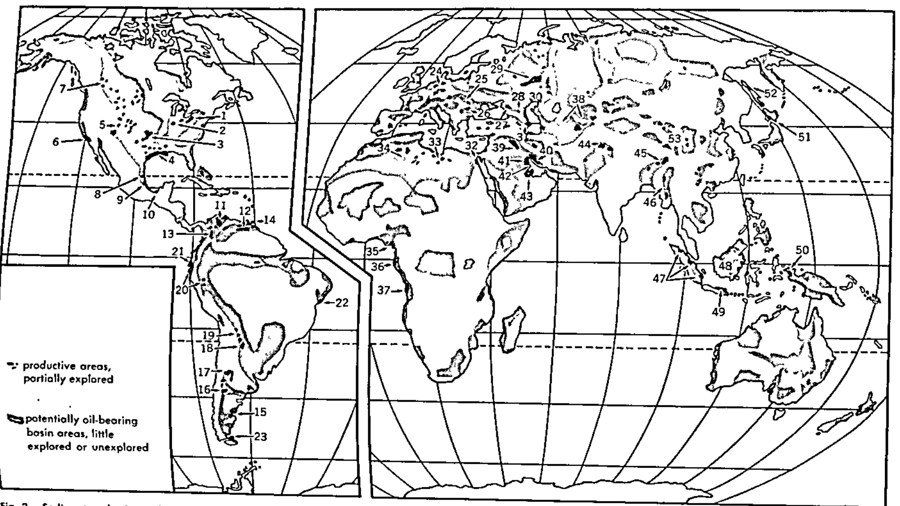


Fig. 2. Sedimentary basins and outstanding oil field areas. Numbers identify area names listed in Table 2.

produce as much additional oil as was recovered by primary methods, which ordinarily averages only about one third of reservoir oil. The largest secondary recoveries are to be expected in older fields; future production research and experience are likely to extract a greater proportion of oil from newer fields by primary recovery. Secondary recovery may add a volume of oil equal to 50% and, in a few cases, possibly 80% or more of that available through primary recovery.

Since reserves are normally calculated in terms of current techniques and costs, anything that lowers the costs of finding and development adds to available reserves. It is anticipated that research of various kinds will continue to increase available reserves.

Outstanding oil field areas are listed and identified in Table 2 and on the map of Fig. 2.

Petroleum gas. Commonly referred to as natural gas, this is found in sedimentary basins in four *types of occurrence.* It occurs *in solution in the* crude oil; in the upper part of the reservoir capping the oil; alone as a dry or nearly dry gas, mainly methane; and at considerable depth, as a so-called condensate, which is a gas carrying considerable liquid hydrocarbons in solution. The distribution of natural gas fields is nearly the same as that of oil fields shown within sedimentary basins on the map of Fig. 2.

An average of 5000–6000 ft³ of petroleum gas has been discovered in the United States for every barrel of crude oil. As approximately 6000 ft³ of petroleum gas has the heat value equivalent of a barrel of oil, the importance of such large volumes of gas is apparent. Unfortunately, markets for gas are lacking in many foreign fields, with the result that much of the gas may be burned in flares at the field However, in many fields gas is returned to the reservoir for storage and to help maintain pressures. In the United States, where good and growing markets exist, the gas is normally conserved, and careful estimates of reserves are made. As of January 1, 1957, these were calculated by the American Gas Association at 238×10^{12} ft³, equivalent in heat value to nearly 40×10^{9} barrels of oil. A reasonable estimate places the ultimate potential gas resources of the United States at $1000\text{–}1200 \times 10^{12}$ ft³. If all were used, the volume of total ultimate natural gas resources of the world would be upwards of 6000×10^{12} ft³.

Oil shale. Although it contains little or no true oil, what is termed oil shale is a compact, fine-grained sedimentary rock which contains a solid organic bitumen, called kerogen, intimately mixed with the mineral matter. On heating the shale in retorts to about 900°F, the solid kerogen breaks down to yield an oil resembling petroleum. This product is, however, highly unsaturated; it is thus much like the products resulting from thermal cracking of petroleum. The oil from the retorts may be burned directly as a fuel or synthesized to form gasoline and other products.

Up to the present, the cost of mining the shale and recovering the oil has restricted its use to a few limited areas of the world where petroleum or petroleum products have not been adequately available or where shale oil production is subsidized. [illegible] shales.

The world possesses vast resources of oil shale which stand ready to supply its fuel needs whenever the supplies of crude oil become inadequate. These shales vary greatly in richness, ranging from less than 10 to as much as 50 gals or more of oil *per ton.* As exploitation costs are reduced through research and experience, permitting the use of less and less rich shales, the volumes of shale oil reserves become rapidly greater.

The greatest known potential resources of shale oil occur in the Green River formation of Eocene age in northwestern Colorado, northeastern Utah, and southwestern Wyoming. The richest of these are in a 1400-mi² area known as the Piceance Creek Basin in northwestern Colorado. Here the Green River formation attains a maximum thick-

Table 2. Outstanding oil and gas fields*

Western Hemisphere	Eastern Hemisphere
North America	*Europe*
1 Appalachian	24 German fields
2 Eastern Interior	25 Polish fields
3 Mid-Continent	26 Romanian Ploesti
4 Gulf Coast	27 Maikop, U S S R
5 Rocky Mountain	28 Grozny, U S S R
6 California	29 Ural-Volga
7 Edmonton, Canada	30 Caspian
8 Tampico, Mexico	31 Baku, U S S R
9 Tuxpan, Mexico	
10 Puerto Mexico	*Africa*
	32 Eastern Egypt
South America	33 Libya
11 Maracaibo Basin	34 N E. Nigeria
12 Orinoco Basin	35 Nigeria
13 Colombia fields	36 Gabon and Congo
14 Trinidad	37 Angola
15 Comodoro Rivadavia, Argentina	*Southwestern Asia*
16 Neuquen–So Mendoza, Argentina	38 Various S W Siberia
	39 Iraq
17 N. Mendoza, Argentina	40 Iran
18 Salta, Argentina	41 Kuwait
19 Bolivia	42 Bahrein-Qatar
20 Peru	43 Saudi Arabia
21 Ecuador	
22 Bahia, Brazil	*Southeastern Asia and East Indies*
23 Southern Chile	
	44 West Pakistan
	45 Assam, N E India
	46 Burma
	47 Sumatra
	48 Borneo
	49 Java
	50 New Guinea
	51 Japan
	52 Sakhalin
	53 Various China fields

* Numbers indicate locations on map, Fig 2

ness of 3500 ft. John R. Donnell, *Preliminary Report on Oil Shale Resources of Piceance Creek Basin, N. W. Colorado*, USGS Survey Bull. 1042-H, 1957, reports the following indicated yields:

Indicated yield ($\times 10^9$ *bbl*)	*Average content* (*gal/ton*)
7.5	45
98.8	30
194.0	25
959.0	15

In an official governmental report of 1955, A. C. Rubel estimates that there are some 1000×10^9 barrels of shale oil in the United States in shales averaging 10 gal of oil per ton. Nearly half of these shales are in the east central part of the United States.

It is estimated that upwards of 100×10^9 barrels of shale oil may be considered economically recoverable now or in the near future in the United States, and that improved technology and economic factors may add greatly to shale oil reserves as the need develops.

Abroad, shale oil has been produced on a moderate scale for many years. Scotland, Sweden, Estonia, Manchuria, Germany, and the Union of South Africa have been the principal producers. Reliable estimates of reserves or potential resources are not available. Table 3 lists the countries possessing shale oil reserves and indicates a rough order of magnitude for them.

Shales have been processed experimentally in the United States, Canada, Sicily, Yugoslavia, Bulgaria, and New Zealand. Oil shales are also known in Arabia, Argentina, Bolivia, Burma, Chile, Colombia, Czechoslovakia, Ecuador, England, Poland, India, Ireland, Israel, Italy, Mongolia, Norway, Panama, Peru, Switzerland, Syria, Thailand, Turkey, Uruguay, Venezuela, and Wales.

Table 3. Principal foreign countries possessing shale oi

Country	Resources, $\times 10^9$ bbl		Operations
	Potential	Proved	
Australia*	75	55	Minor and intermit tent
Brazil	100,000		Insignificant
Canada	Large		Best known are Al bert Shales of New Brunswick
Estonia	?	6000	Shales yielding 45–55 gal/ton processed since 1921
France	300	20	1400-bbl/day plant in operation recently
Germany	1700	60	Modest since World War II
Manchuria	2000	?	In World War II at tained 3600 bbl/day
New Zealand	200		
Scotland	180	40	Long continuous op eration on shales yielding as low as 15 gal/ton
Spain	?	80	Very modest
Sweden	4000	10	Long continuous on moderate scale
Union S. Africa*	?	6	Long continuous on small scale
U.S.S.R.	7000	?	Extent of operations unknown

* The bitumen occurs as solid hydrocarbons, called tor banites, rather than as oil shales, with yields of from 50 to 100 or more gallons of oil per ton

It may be said that commercially proved world reserves of shale oil are measurable in scores of billions of barrels, but probably not over 150,000,000,000 bbl. On the other hand, potential shale oil resources of the world are enormous, measurable in trillions of barrels. See OIL SHALE.

Oil sands. In addition to bitumen-bearing shales, vast hydrocarbon accumulations occur as bitumen or heavy oil in sands, the so-called tar sands. The

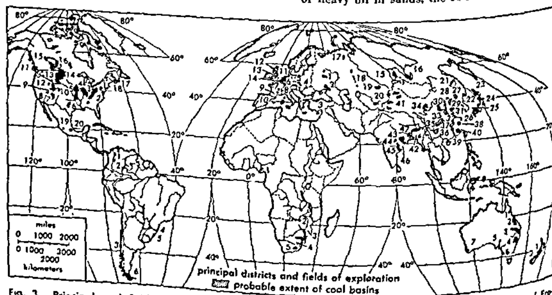


Fig. 3. Principal coal fields and probable extent of outstanding coal basins of the world. (From N. A. Bengtson and W. Van Royen, *Fundamentals of Economic Geography*, 4th ed., Prentice-Hall, 1956)

most extensive known is the Athabaska tar sand accumulation in Alberta, western Canada. The sands cover an area of 12,000 mi^2. The oil content is estimated at 300×10^9 barrels. Some areas of the sands contain as much as 100,000,000–200,000,000 barrels of oil per square mile. Perhaps $5–10 \times 10^9$ barrels of this oil can be made commercially recoverable in the near future.

In a 1952 report to the National Petroleum Council, A. C. Rubel estimated the recoverable oil content of tar sands in the state of Utah at over 600,000,000 barrels and that of five tar-sand accumulations in California at 160,000,000 barrels.

Tar sands occur in huge aggregate volumes in Venezuela, the Middle East, and also in Brazil, Colombia, Ecuador, and apparently in other less well known areas of the world. A large deposit of tar or asphalt has long been mined in Trinidad. Total known potential tar-sand oil resources of the world are now estimated to be of the order of some 750×10^9 barrels. Undoubtedly the known volumes of both tar sands and oil shales will be multiplied by future discoveries. Little of this oil is economically extractable today, but it remains as a potential resource as need arises. See OIL SAND.

Synthetic liquid fuels from coal. The volume of synthetic liquid fuel which can be manufactured from coal appears to be comparable to that available from all other sources combined. As one example, deposits of lignite and subbituminous coal in western United States and Canada, not now salable for commercial use in distant markets, could be hydrogenated or otherwise converted to more than 1×10^{12} barrels of liquid petroleum. However, present indications are that the economic production of petroleum from shales and tar sands will be marketed well ahead of synthetic fuels from coals and lignites. *See* FISCHER-TROPSCH PROCESS.

Coals. By far the most abundant of the fossil fuels are the coals. The world-wide distribution of coal field regions and the probable extent of some outstanding coal basins are indicated on the map of Fig. 3. Numbers on the map identify the coal fields as listed in Table 4. The sketch map of Fig. 4 shows graphically the concentration of production in the United States and a few countries of western and central Europe, such as the United Kingdom, France, Belgium, Netherlands, Germany, Poland, and Czechoslovakia. This map also shows evidence of the rising coal productions in U.S.S.R., Japan, China and Manchuria, and India. More recent data on coal production are given in Table 5. Table 6 shows six coals in metamorphic sequence; for a more technical consideration, *see* COAL.

Table 4. Major coal fields of the world*

North America

1 Pennsylvania Anthracite region
2 Appalachian region
3 Northern Interior region
4 Eastern Interior region
5 Western Interior region
6 Southwestern Interior region
7 San Juan River region
8 Southwestern Utah region
9 Uinta region
10 Denver region
11 Green River region
12 Hams Fork region
13 Bighorn Basin region
14 Fort Union region
15 Peace River field
16 Alberta fields
17 Sydney field
18 Pictou field
19 Michoacan field
20 Puebla field

South America

1 Rio Pamplonita field
2 Santander field
3 Arauco field
4 Santa Catarina field
5 Porto Alegre field
6 Punta Arenas field

Australia

1 Mt Mulligan field
2 Mackenzie–Dawson River field
3 Ipswich-Tawbo field
4 Newcastle field
5 Leigh Creek field
6 Wonthaggi field
7 Collie field

New Zealand

1 Westport field
2 Greymouth field

Africa

1 Enugu field
2 Mid-Zambesi–Warkie field
3 Limpopo-Lebombo field
4 Natal Zululand field
5 Cape Province field

Europe and Asia

1 Moscow field
2 Donets field
3 Zongouldack field
4 Upper Silesian basin
5 Central Germany brown coal fields
6 Ruhr Coal basin
7 North France–Central Belgium–Netherlands field
8 Saar basin
9 Central Plateau fields
10 Oviedo field
11 Northumberland-Durham field
12 Scottish Coal fields
13 Yorkshire fields
14 Welsh fields
15 Kuznetsk field
16 Irkutsk field
17 Pechora field
18 Ural fields
19 Karaganda field
20 Uzbek field
21 Talaya field
22 Vladivostok field
23 Ishikari field
24 Fukushima-Ibaragi field
25 Yamaguchi field
26 Fukuoka field
27 Southern Liaoning region
28 Central Jehol region
29 Northeast Hopei region
30 Shansi-Shensi region
31 Central Shantung region
32 N. Kiangsu–S Shantung region
33 N. Honan region
34 W. Kansu region
35 E. Szechuan region
36 E. Yunnan region
37 E. Hunan–W. Kiangsi region
38 S. W. Chekiang region
39 S. E. Hunan region
40 S. Fukien region
41 Sinkiang region
42 E. Himalayas, Assam, and Bengal fields
43 Central India field
44 Satpura basin
45 Hyderabad field
46 Orissa field
47 Bihar field

* Based on a table in N. A. Bengtson and W. Van Royen, *Fundamentals of Economic Geography*, 4th ed., Prentice-Hall, 1956.

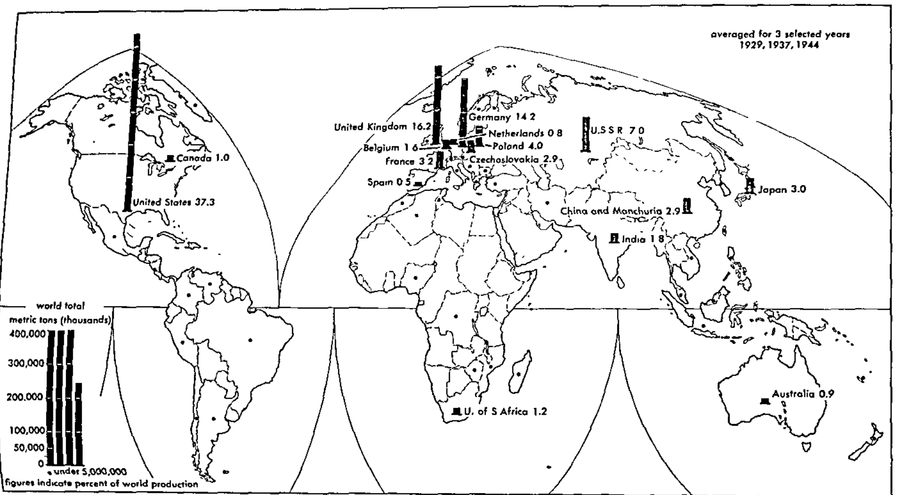


Fig. 4. Sketch map of world production of coals by countries. Lignite is considered in coal equivalent. (From O. Bowles and W. Van Royen, *Atlas of the World's Resources 1952–1954*, vol. 2, Prentice-Hall, 1952)

Table 5. World coal production, by countries and grades, 1956–1958, in thousands of short tons

Country	1956				1957				1958[a]			
	Anthrac.	Bitum.	Lignite	Total	Anthrac	Bitum	Lignite	Total	Anthrac.	Bitum.	Lignite	Total
Canada		12,574	2,342	14,916		10,940	2,249	13,189		9,434	2,253	11,687
Greenland		8		8		19		19		19		19
Mexico		1,332		1,553		1,566		1,566		1,621		1,621
United States	28,900	497,997	2,575	523,775	23,338	490,097	2,607	516,042	21,171	408,019	2,427	431,617
North America	28,900	512,131	5,230	548,231	25,338	502,622	4,856	532,838	21,171	419,093	4,680	444,944
Argentina		169		169		230		230		285		285
Brazil[c]		2,463		2,463		2,285		2,285		2,429		2,429
Chile		2,511		2,511		2,310		2,310		2,011		2,011
Colombia		2,094		2,094		2,480[b]		2,480		2,425		2,425
Peru[d]		160		160		155		155		192		192
Venezuela		34		34		39		39		40		40
South America		7,431		7,431		7,499		7,499		7,381		7,381
Algeria[d]		327		327		260		260		169		169
Belgian Congo		453		453		477		477		324		324
Madagascar						1		1				
Morocco	531			531	574			574	562			562
Mozambique		240		240		298		298		273		273
Nigeria		852		852		913		913		1,036		1,036
Rhodesia, Nyasaland		3,918		3,918		4,247		4,247		3,897		3,897
Tanganyika		1		1		1		1		1		1
Union South Africa[d]		37,040		37,040		38,325		38,325		40,879		40,879
Africa	531	42,871		43,402	574	44,522		45,096	562	46,579		47,141
Australia		21,587	11,827	33,414		22,310	12,030	34,340		22,844	13,041	35,885
New Zealand[c]		897	2,046	2,943		931	1,994	2,925		939	2,108	3,047
Oceania		22,484	13,873	36,357		23,241	14,024	37,265		23,783	15,149	38,932
Albania			255[b]	255			260	260			282	282
Austria		183	7,419	7,602		168	7,581	7,749		155	7,158	7,313
Belgium[d]		32,475		32,475		32,062		32,062		29,831		29,831
Bulgaria	137		11,787[c]	11,924	150[b]		12,937[c]	13,107	150[b]		13,867[c]	14,017
Czechoslovakia		25,806	51,036	76,842		26,655	55,235	82,890		29,321	57,100	86,421
Denmark			1,534	1,534			2,100[b]	2,100			3,300[b]	3,300
France		60,773[d]	2,484	63,257		62,606[d]	2,528	65,124		63,626[d]	2,555	66,181
West Germany[f]		151,406[d]	104,976	256,382		150,118[d]	106,716	256,832		149,196[d]	103,052	252,248
East Germany		3,024[d]	226,928	229,952		3,035[d]	234,346	237,381		3,200[d]	236,961	240,161
Greece			880	880			1,100	1,100			1,077	1,077
Hungary		2,619	20,080	22,699		2,510	20,858	23,368		2,895	23,826	26,721
Ireland		239[d]		239		278[d]		278		285[d]		285
Italy		1,188[d]	445	1,633		1,128[d]	425	1,553		793[d]	898	1,698
Netherlands		13,047	298	13,345		12,540	317	12,857		13,095	281	13,376
Poland		104,884	6,816	111,700		103,723	6,563	110,286		104,699	8,313	113,012
Portugal		456[d]	161	617		550[d]	203	753		625[d]	172	797
Rumania[e]		210[d]	6,900	7,110		275[d]	7,400	7,675		330[d]	7,800	8,130
Saar		18,838		18,838		18,139		18,139		18,103		18,103
Spain		14,165[d]	2,125	16,290		15,356[d]	2,777	18,133		15,907[d]	2,927	18,834
Spitsbergen[g]		816		816		858		858		760[b]		760
Sweden		324		324		335		335		332		332
Switzerland[h]		11		11		11		11		11		11
United Kingdom		248,646[d]		248,646		250,464[d]		250,464		241,721[d]		241,721
Yugoslavia		1,258	17,493	18,851		1,353	18,497	19,850		1,332	20,928	22,260
Europe	137	680,468	461,617	1,142,222	150	682,162	480,861	1,163,173	150	676,222	490,497	1,166,869
U.S.S.R.[j]		334,772[d]	138,340	473,112		360,455[d]	149,914	510,369		385,809[d]	160,937	546,746
Afghanistan		26		26		30		30		22[b]		22
China[b]		116,700		116,700		144,100		144,100		297,600		297,600
India		43,994		43,994		48,720		48,720		50,777		50,777
Indonesia		914		914		788		788		661		661
Iran[l]		209		209		194		194		187		187
Japan		51,318[d]	1,676	52,994		57,025[d]	1,832	58,857		54,300[d]	1,744	56,044
Republic Korea	2,001			2,001	2,691			2,691	2,944			2,944
North Korea[m]	4,500			4,500	5,500			5,500	7,600			7,600
Malaya		204		204		171		171		75		75
Pakistan		722		722		578		578		689		689
Philippines		168		168		211		211		119		119
Taiwan		2,788		2,788		3,214		3,214		3,508		3,508
Thailand			95	95			110	110			138	138
Turkey		6,490	3,318	9,808		6,917	3,899	10,816		7,234	4,212	11,446
North Vietnam	1,213			1,213	1,200[b]			1,200	1,200[b]			1,200
South Vietnam	2			2	13			13	22			22
Asia[i]	7,716	223,533	5,090	236,339	9,404	261,948	5,841	277,193	11,766	415,252	6,074	433,012
Other countries[k]				110				110				110
World total[h]	37,284	1,823,690	624,140	2,485,224	35,466	1,882,449	655,498	2,573,321	33,649	1,974,019	677,357	2,683,133

SOURCE: Compiled from figures supplied by U.S. Bureau of Mines. [a] Preliminary. [b] Estimate. [c] Includes lignite. [d] Includes anthracite. [e] Includes bituminous. [f] Included under bituminous are about 2000 tons of Pechelbron coal in each of the three years. [g] Includes 368, 434, and an estimated 440 tons produced from the U.S.S.R.-controlled mines in the years 1956, 1957, and 1958 respectively. [h] Estimated total includes also anthracite and lignite. [i] Does not include U.S.S.R. [j] Includes Sakhalin. [k] Total includes also anthracite and lignite. [l] For year ending following March 20. [m] Estimated total also includes lignite. [n] Does not include Siberia and Sakhalin, which are included with U.S.S.R.

Table 6. Six classes of coal in metamorphic sequence from plant origin to anthracite*

0 Woody tissue (lignocelluloses) and other plant materials 1. Peat (Quaternary) 2 Brown coals—less than 8300 Btu/lb (Tertiary) 3 *Laminated lignites*—less than 8300 Btu/lb (chiefly Tertiary) 4 Subbituminous coals—8300–13,000 Btu/lb (chiefly Cretaceous) Subbituminous C Subbituminous B Subbituminous A	5 Bituminous coals—more than 11,000 Btu/lb (*Mesozoic, Carboniferous*) High volatile C High volatile B High volatile A Medium volatile Low volatile 6 Anthracite coals—less than 14% volatiles and 86% fixed carbon; Btu is *generally lower than bituminous* because of very low volatiles (Carboniferous) Semianthracite Anthracite Metaanthracite

* Modified from a *classification* in P C Putnam, *Energy in the Future*, Van Nostrand, 1953 This classification conforms with standards of U S Bureau of Mines

Northern Hemisphere concentrations. Figures 5, 6, 7, and 8 present somewhat greater detail on several aspects of coal and coal fields in the United States and Canada. Great Britain. Northwestern Europe, U.S.S R., and Southern and Eastern Asia. Within these *Northern Hemisphere* areas are fully 96% of the estimated coal reserves of the world. South America, *Africa, Australia–New Zealand*, and Oceania have relatively little coal.

Coal reserves. It is difficult to determine the size of proved or potential coal reserves. *There is probably* more coal in the world than man can ever use, *but nothing* is a reserve or potential reserve that cannot one day be economically *produced* and utilized. Therefore, what constitutes reserves depends ultimately *on how far* into the total supply it will be economical *for man* to delve. Because of the widely different interpretations applied to economic factors and the difficulty of defining the terms which should *control* estimates, published figures of coal reserves have varied widely.

The origin of coal deposits, unlike the more hidden origin and occurrence of petroleum, may be witnessed today *and coal seams* may be observed or measured at *the earth's* surface. It may be partly for these reasons that many of the early and carefully studied estimates of *gross coal* reserves have stood comparatively well with the *passage* of time. The criticisms that have arisen have stemmed largely from misinterpreta[illegible] been ignored or *overlooked*. This may be illustrated by considering the history and present status of United States *coal* reserve estimates, and then those of the rest of the world.

United States reserve problems. The United States has about 300,000 mi² of known coal-bearing area. In 1909, M. R. Campbell stated that the United States has 3.1×10^{12} tons of coal of all ranks in seams over 14 in. thick within 3000 ft of the surface and with no more than 30% ash. He pointed out, however, that much of this coal lies too deep or is in seams too thin to be economically mineable by methods then in use. Allowing for these two factors, Campbell arrived at a figure of 1.4×10^{12} tons. He went further to point out that over one-third of this coal is of low rank and not readily utilizable. However, despite the clarity of Campbell's statements, the figure of 31×10^{12} continued to be used, even in official governmental statements as late as 1944. Similarly large or larger estimates were made later.

In recent years there has been a gradual acceptance of the concept that economically recoverable reserves in the United States are only a fraction of the 3×10^{12} ton estimate widely used for about four decades. Many have contributed various factors of correction. P. C. Putnam in 1953 applied five different factors to arrive at a potential coal reserve for the United States of 310×10^9 or one-tenth that of Campbell's gross estimate. The U.S. Geological Survey estimated remaining coal reserves as of January 1, 1953, at 1.9×10^{12} short tons within certain specified limits of inclusion, but reduced this on the basis of 50% recovery in mining to 0.95×10^{12} tons (7×10^9 anthracite, 525×10^9 bituminous, 186×10^9 subbituminous, 232×10^9 lignitic coals). Only 237×10^9 tons, however, was considered recoverable by current production methods—one-eighth the 1.9×10^9 figure. Table 7 summarizes their conclusions.

Table 7. USGS estimates of United States coal reserve distribution, January 1, 1953*

Area	Anthracite	Bituminous	Subbituminous	Lignite	Tot
Eastern	6.3	223			229.
Interior	.1	206			206
Western	.4	116	136	119	371
United States totals	6 8	545	136	119	806

* Distribution in net tons ×10⁹, equivalent to 13 000 Btu coal.

In May, 1957, Ivan A. Given arrived at the figure of 152×10^9 tons of coal reserves east of the Mississippi River in seams mineable at or slightly above today's prices. Allowing one-third for losses in mining and preparation, Given brings the total down to 100×10^9 tons. He concludes that "all the coal anyone will need will be available at reasonable prices for a long time to come in the area east of the Mississippi."

Canadian reserves. The coal reserves estimates of Canada, long spoken of as being upwards of 1.2×10^{12} tons of mostly low-rank coal, have also been reduced in more recent years to about one-tenth the early gross figures. The estimate may be further reduced because the western interior location of reserves is distant from the principal markets of eastern Canada.

Fig. 5. Principal coal fields of the United States and Canada distinguished as to location, extent, and principal types of coal. (From V. C. Finch, G. T. Trewartha, A. H. Robinson, and E. H. Hammond, Elements of Geography, 4th ed., McGraw-Hill, 1957)

World-wide reserve patterns. Attempts to arrive at estimates of other foreign reserves entail possibilities for error that, for some countries, are likely to be greater than in the case of the United States and Canada. D. C. Jones and J. W. Hunt in 1949 presented the world reserves data shown in Tables 8 and 9. Approximately the same figures, converted to short tons, had been reported to the Twelfth International Geological Congress in 1913. A complete assessment of the world's coal re-

Table 8. Coal resources of the world*

Area	Anthracite and some dry coals	Bituminous coals	Subbituminous and lignitic coals	Totals
North America	21,842	2,239,683	2,811,906	5,073,431
South America	700	31,397		32,097
Europe	54,336	693,162	36,682	784,180
Asia	407,637	760,098	111,851	1,279,586
Africa	11,662	45,123	1,054	57,839
Oceania	659	133,481	36,270	170,410
Totals	496,846	3,902,944	2,997,763	7,397,553

* In millions of metric tons (1.1023 short tons). Estimated supplies of coal in beds 12 in. or more in thickness and with less than 4000 ft of cover, and in beds 2 ft or more thick lying between 4000 and 6000 ft below the surface. The figures of this table are generally considered unduly optimistic.

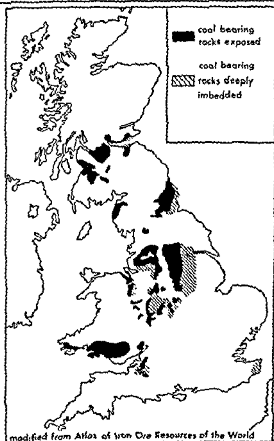


Fig. 6. Coal fields of Great Britain.

Table 9. Coal resources by countries*

Country	Metric tons × 10^9	Area of world or continent, %	Remarks
North America	5073	68 6	
Canada	1234	24.3	77% low-rank coal
United States	3839	75.6	Over half bituminous or higher-rank
Balance	0.1	0.1	Alaska, 19 × 10^9 tons; mostly low-rank
South America	32	0.4	
Colombia	27	84.4	Mostly bituminous; thin beds
Peru	2	6.3	High-rank; good thicknesses
Chile	2	6.2	Mostly low-rank
Brazil and others	1	3.1	Largely low-grade bituminous
Europe	784	10.6	
Germany	422	54.0	High-rank; Europe's leading producer
Great Britain	190	24.0	High-rank, once Europe's leading producer
Austria	77	9.8	Two-thirds bituminous, balance low-rank
U S S R.	60	7.7	94% anthracite and bituminous
France	17	2.2	Bituminous or higher-rank
Belgium	11	1.4	Bituminous or higher-rank
Balance	7	0.9	
Asia	1280	17.3	
China	998	78.0	Mostly bituminous and anthracite
Siberia	174	13.5	Data very inadequate
India	79	6.2	Mostly high-rank
Indochina	20	1.6	Mostly high-rank
Japan	8	0.6	
Balance	1	0.1	
Africa	58	0.8	
Union S. Africa	56	97.0	Largely Permian of medium rank
Balance	2	3 0	
Oceania	170	2.3	
Australia	167	98.0	Mostly Permian of bituminous rank
Balance	3	2.0	Nearly all in New Zealand
World	7397	100.0	

* In billions of metric tons (1.1023 short tons). Estimated supplies of coal in beds 12 in. or more in thickness and with less than 4000 ft of cover, and in beds 2 ft or more thick lying between 4000 and 6000 ft below the surface. The amounts of metric tons are generally considered unduly optimistic.

sources was the main business of that Congress. Again, it should be pointed out that it is the gross *volume figures* that have been generally cited *although the* estimates were classified as actual, probable, and possible.

Later revisions of world estimates have been largely *concerned, as in* the United States, with what proportion of the statistical totals is economically *utilizable. With* one notable exception, the U.S.S.R., there has been a drastic reduction. Significantly, *for most countries* the revisions have resulted in present estimates comparable to the "actual" reserves as set in 1913.

P. C. Putnam summarized recent views on the extent of economically recoverable reserves of the world in terms of heat content expressed in quanta Q which equals 10^{18} Btu or 38×10^9 *tons of bituminous* coal. Average Btu/lb heat content of *United States* coals is given as 12,700 for anthracite, 13,100 for bituminous, and 8000 (*weighted average* based on proportionate volumes) for subbi- [illegible] Table 10 [illegible] t content [illegible] coverable by present methods at substantially *present costs. To* these are added figures representing equivalent tonnages of bituminous coal.

Based on experience with reserve estimation of many kinds, it seems likely that for many countries actual recoveries will prove to be *considerably* greater than now forecast. This may or *may not* hold true in the case of the U.S.S.R. There the estimated volume of economically recoverable coal arrived at by Putnam, instead of being *reduced to* one-fifth or one-tenth of the gross figures reported to the Twelfth International Geological *Congress* in 1913, is increased by 50%. No doubt the early gross figures were too low due to *inadequate* exploration. It may be of *interest to* add that Put-

Table 10. Estimates of heat content and equivalent tonnages of economically recoverable coal reserves of the world, 1950*

Country	Heat content, Q	Equivalent in tons bituminous coal, × 10^9
United States	6	228
Canada	2	76
United Kingdom	1	38
Other Free-World countries	7	266
China	6	228
U.S S.R.	10	380
	32	1216

* Table adapted from P. C. Putnam, *Energy in the Future*, Van Nostrand, 1953

Fig. 7. (a) Principal coal fields of Northwestern Europe. (b) Principal coal and petroleum fields of U.S.S.R. (From V. C. Finch, G. T. Trewartha, A. H. Robin and E. H. Hammond, *Elements of Geography*, 4th ed. McGraw-Hill, 1957.)

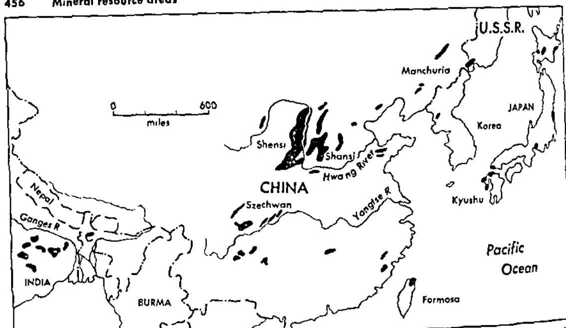


Fig. 8 Principal coal fields of Southern and Eastern Asia. (*From V. C. Finch, G. T. Trewartha, A. H. Robinson, and E. H. Hammond, Elements of Geography, 4th ed., McGraw-Hill, 1957*)

nam's 1953 figure of 10 Q(380 × 10^9 tons) was about one-fourth that claimed at the time by the U.S.S.R. *See* COKE; ENERGY SOURCES. [L.G.W.]

Bibliography: E. Ayres and C. A. Scarlott, *Energy Sources—the Wealth of the World*, 1952; P. C. Putnam, *Energy in the Future*, 1953; U.S. Dept. of State, *Energy Resources of the World*, Publ. 3428, 1950; L. G. Weeks, Concerning estimates of potential oil reserves, *Bull. Am. Assoc. Petrol. Geologists*, 34(10):1947–1953, 1950; L. G. Weeks, Factors of sedimentary basin development that control oil occurrence, *Bull. Am. Assoc. Petrol. Geologists*, 36(11):2071–2124, 1952; L. G. Weeks, Fuel reserves of the future, *Bull. Am. Assoc. Petrol. Geologists*, 42(2):431–438, 1958. L. G. Weeks, The next hundred years energy demand and sources of supply, *Bull. Am. Assoc. Petrol. Geologists*, 1960.

Mineral resource areas

Regions characterized by either a useful mineral

glass, one-third of its land surface has no economic minerals, another third is discouragingly deficient, and about one-third is potential mineral resource region with resources differing from domain to domain. Discontinuity between areas can be attributed partly to absence of genetic influences causing mineral segregation; however, diversity from domain to domain, each subject to the same genetic conditions, indicates fundamental diversity in element distribution throughout the earth. Areas marked by unusual abundance of an element or mineral are metallographic provinces.

Origin. Mineral resource areas originate partly through the existence of a material in slightly abnormal abundance, and partly by the segregation

concentrations by natural

strik

a, Co-

lombia, where beryllium is abundant and practically every type of rock reorganization produces the variety of beryl known as emerald. In fact, within this region, emerald has been found in every way minerals have been known to occur, most of them unusual for beryl. The genetic processes that segregated the beryllium to crystallize into this mineral acted upon a body of rock with hot spring solutions, reorganized it at high temperature, and removed the rejected beryllium to crystallize in fissures or rocks of the peripheral zone.

Segregation of elements or minerals occurs by various means. Mechanical processes of selective removal leave enriched residues of refractory heavy minerals at the surface, and water wash produces concentrations in river gravels near the depth of scour, and streaks on beaches at the uprush zone. Surface chemical solutions often remove one component to leave a residue and elsewhere permit one component to crystallize selectively. Tropical residues are enriched in aluminum or iron hydrate or iron and manganese oxide (Fig. 1). Temperate or volcanic weathering removes iron from the soil, and bacterial action or neutralization by alkaline sea water precipitates it to form iron deposits. Evaporation of sea water in the trade-wind and horse-latitude deserts leaves stratiform salt residues.

Aqueous underground solutions acquire a variety of components and carry these through porous strata, lava flows, fault fractures or shears, disrupted layered rocks, or simply a joint system. Combinations of elements separate out as individual minerals or mineral assemblages where solu-

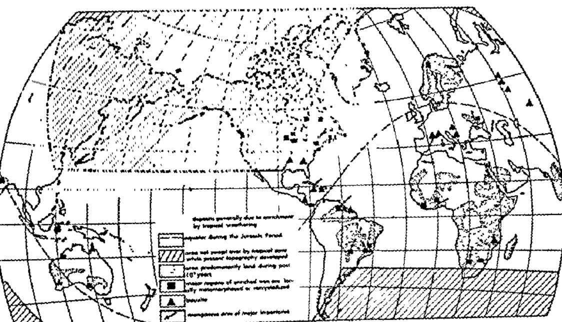


Fig. 1. Resource regions within area swept over by tropical-type climate since present topography began to develop.

bility is exceeded. Acquisition of elements results as often from accentuated ionization as it does from increased temperature or pressure. The latter may be promoted by surface processes, but more frequently is abetted by earth movements permitting ascent of heating agents from deep within the earth. Ionized solutions rework great earth masses underneath unstable zones, and by selective crystallization segregate the unwanted metallogenic elements from the petrogenic ones, which remain as rock bodies and magmas. Selective crystallization from solution attends decrease in pressure or temperature or ionization, usually by reduction of acid solutions and by oxidation of alkaline solutions, and also as a result of other ionic disturbances. Surface-enriched copper deposits are examples of precipitation through solution reduction, and most ores of lead and zinc and the chalcopyrite-bornite deposits of copper are attributable to solution oxidation.

Locally the solutions merely transform a mineral already present (serpentine changed to talc) or they may move and recrystallize the mineral (massive serpentine redeposited as chrysotile asbestos). Magmas are silicate solutions; crystallization of their petrogenic components segregates into a liquid residue the metallogenic elements of the region that was worked over to make the magma (titaniferous magnetite bodies and some platinum occurrences in peridotite).

Geologic associations. Recurrent weathering enriches iron ores and, in consequence, recurrently elevated regions of ancient rocks, called shields, have the largest iron deposits (Fig. 1). Copper, lead, and zinc deposits are principally in lode-type occurrences left in the rock openings produced by mountain-making movements, and appear where copper provinces (or lead and zinc provinces) became involved in orogenic disturbances (Fig. 3).

Metallographic provinces for copper (Fig. 2) have defined limits and represent local abnormal abundance for this element. Provinces for tin, cobalt, uranium, and silver show a distribution that is distinctively different from copper. Some areas such as the New England states and Ecuador lie between phenomenal mineral resource areas, and are themselves resource deserts for all metals.

Only limited geologic regions had genetic processes that operated close enough to the present surface of the earth to place the resulting segregations within reach of existing exploration methods and mining procedures. Weathering enrichments appear principally where the rocks have been swept over by tropical climates since the present topography began developing. Iron must be enriched so much in such large volume in order to constitute ore that recurrent enrichments are essential, and most great deposits are found where the region has persisted as a land area for eras of time (Fig. 1). Base metals, precious metals and ferroalloy elements are found principally in regions with mountain-type structure (Fig. 3). Salts, phosphorites, and other stratiform deposits appear in plains and plateau regions of almost undeformed strata (Fig. 4).

Categories of materials. Mineral resources are sometimes used for the mineral's own properties, as in the instance of kaolinite for ceramics (industrial mineral group), whereas others are recovered for extraction of the desirable element, like the

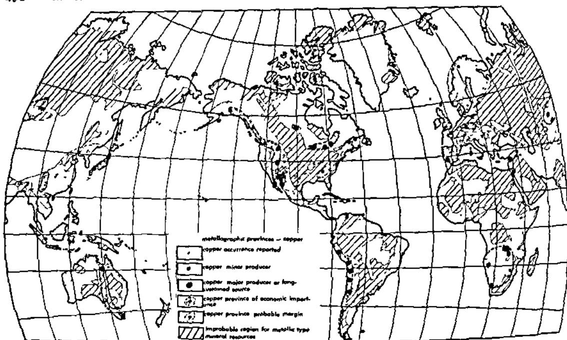


Fig. 2. Copper metallographic provinces of the world.

aluminum in bauxite (ore mineral group). Significant industrial minerals include the ceramic minerals such as nepheline (igneous), kaolinite (a ubiquitous mineral), and kyanite, sillimanite, andalusite, and dumortierite, which are metamorphosed clay derivatives. Muscovite is widespread but occurs in commercial quantity only in pegmatites. Diamonds are from alluvial deposits concentrated during erosion of either kimberlite (South Africa, Tanganyika, Belgian Congo, Brazil, Yakutia, and Arkansas) or eclogite (Gold Coast, India, Venezuela, and the Guianas). Asbestos is the result of solution and recrystallization of serpentine in Transvaal, Swaziland and adjacent Southern Rhodesia, Ural Mts., Norway, Quebec, and Vermont. Agricultural chemicals and common salt are segregated along seacoasts under contemporary tradewind deserts or horse-latitude high-pressure zones. Phosphates form extensive beds in Permian strata of Idaho and adjacent states, in later Mesozoic strata of the Moscow basin (U.S.S.R.), the upper Cenozoic strata of Florida, and the elevated reefs of the tropical Pacific Ocean. Common salt is widespread in strata of Silurian, Permian, and Oligo-



Fig. 3. Types of host regions for mineral resource areas.

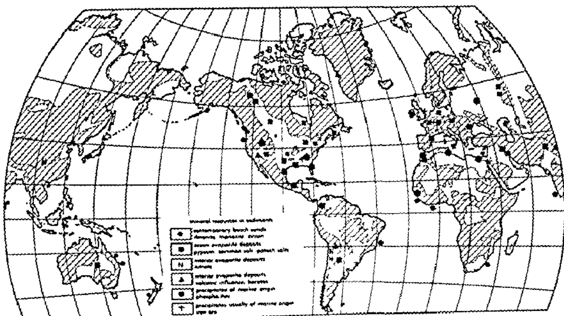


Fig 4 Resource regions for stratiform deposits

cene age, whereas potash salts are obtained from the Devonian strata of the Canadian prairies, the Permian beds near Carlsbad, New Mexico, in Central Europe, the Perm area of Russia, and the Oligocene-age strata of Mulhausen, Germany, and Barcelona, Spain. Elemental sulfur is the principal source of this mineral, but it is also made by natural reduction of sulfates using the energy of petroleum at slightly elevated temperature as the reducing medium. Principal occurrences are on salt domes around the Gulf of Mexico in Louisiana, Texas, Tehuantepec (Mexico), Sicily, Caspian Sea and Volga area (U.S.S.R.), and Rumania.

Enriched iron ores, in deposits exceeding 1×10^9 tons, are in the very old land areas like Minas Gerais (Brazil), Estado Bolivar (Venezuela), Minnesota, Quebec, Transvaal (South Africa), north edge of the southern peninsula of India, Krivoi Rog (U.S.S.R.), and Scandinavia. Lower-content ores in large bodies are in bedded strata of the Appalachians, Newfoundland, and the Jurassic strata of England, France, and Luxembourg. All bauxites were made by tropical weathering during periods of prolonged chemical denudation, as during late Mesozoic weathering on the flanks of the Ural Mts., in Hungary, southern France, the Guianas, Arkansas, and Australia, and by Cenozoic-age weathering in Australia, Indonesia, Belgian Congo, Nyasaland, Brazil, and the Caribbean and Central American region. Manganese markets were dominated for years by production from weathering enrichments at Nikopol in the Ukraine and Tchiatoura in Georgia (U.S.S.R.), but equally large and somewhat similar ores are in Minas Geraes and southern Mato Grosso (Brazil), British Guiana, south of the High Atlas Mts. (Morocco), the Gold Coast, Postmasburg (South Africa), and Bala Ghat (India).

Three great copper provinces dominate the world supply. The largest producer supplies about 35% of the world's needs and extends from Alaska through British Columbia, Montana, Utah, Nevada, Arizona and New Mexico to central Sonora of Mexico. Age of the deposits varies almost 10^9 years. Another large producer, and probably the largest resource, extends from central Peru southward through the Andes into southern Chile, and a third of nearly equal size lies along the Rhodesia-Congo border where deposits are in fractured Mines Series strata. Lesser copper resource provinces are in northern Manitoba, from Michigan to central northern Quebec, and the Kirghiz Steppes from Karaganda eastward to the Yenisei River. The unique stratiform mineralization of the Mansfeld shale is an important but low-grade occurrence in central Europe, while Boliden (Sweden) and Rio Tinto (Spain) had some considerable output in the past. Australia had only the small, low-content deposits in Tasmania until the large new one adjacent to the lead zinc deposits of Mt. Isa was opened.

Titanium, zirconium, and the rare earths come principally from tropical beach sands where the minerals rutile, ilmenite, zircon and monazite are concentrated at the wave uprush line from cubic miles of rock reworked by rivers and the sea. Principal deposits are along the northeast coast of Australia, the Travancore coast of India, the Gambia and Guinea coast of Africa, and the Espirito Santo coast of Brazil; lesser deposits are in the river alluvials of Malaya and Indonesia and the Snake River of Idaho. Some columbium and tantalum appear in large pegmatites and their weathered residues in Western Australia, Nigeria, the eastern Belgian Congo, northern New Mexico, and diagonally southwestward across Brazil, but principally

in cone sheet carbonatites in the Monteregian zone from Montreal to North Bay and northern Ungava (Canada), Southern Rhodesia, Tanganyika, and Uganda, and a sheetlike deposit in the Khibine Tundra of the U.S.S.R. Existence in many other provinces is possible and the short time during which these elements have been sought suggests that this is only a minor part of their total.

The four great occurrences for uranium are (1) in river alluvials, formed before a land flora provided fermentation materials to make soil, as at Blind River (Ontario), the Witwatersrand (South Africa), and Bahia (Brazil); (2) of equal magnitude but of discouragingly refractory nature, the early Paleozoic black bituminous shales of Sweden and the U.S.S.R., the Ohio River basin, and the Phosphoria formation of Idaho; (3) lode deposits which had great original importance and included the entire Haut Katanga (Belgian Congo) and the western edge of the Canadian Shield; (4) the ground-water deposits in and along river channels in the Colorado Plateau country and the Kokand Basin of Uzbekistan.

Production problems. Development of many mineral resource areas has been delayed by problems connected with recovery of the desirable mineral. Costly transportation, inadequate water supply, and limited labor have postponed exploitation. Separation of the minute galena and sphalerite grains from one another and from waste minerals in the ore at Kimberley, B.C., and Mt. Isa, Australia, delayed development of these two great ore districts for a decade. Transportation cost to bring lead, zinc, and copper from the huge deposits on the Rhodesia-Congo border and from Mt. Isa reduces greatly the value of these outstanding contributors to world metal supply. The Peru-Chile copper region is a phenomenal metal domain where water deficiency inhibits expansion; the El Salvador deposit (Chile) could not be developed until Potrerillos was exhausted; farther north, no new mines can be opened and Chuquicamata production cannot be expanded significantly because all water is already being used efficiently.

Minerals are an exhaustible resource and are not regenerated. Mines should be opened with minimum disturbance of human activity. The Conquistadores uprooted the Amerindians from their farms to become miners and left decimating privation when the principal silver and gold was extracted. The development of manganese and other metals in areas of marginal carrying capacity, like Morocco, under French encouragement, kept production below potential and left most people on their land, with mines furnishing merely a supplemental income. The Witwatersrand mines (South Africa) are planned for a minimum life of 50 years so that labor will not have to be resettled too extensively when the deposits are exhausted. No mineral resource area can be exploited recklessly without incurring future hardship at the time of exhaustion. [C.W.B.]

Bibliography: G. W. Bain, *Principles of Geology*, 1959; A. M. Bateman, *Economic Mineral Deposits*, 2d ed., 1950; A. M. Bateman, *The Formation of Mineral Deposits*, 1951; T. S. Lovering, *Minerals in World Affairs*, 1943.

Mineral resources conservation

Conservationists study the supply and use of minerals and ways of satisfying the growing volume and complexity of human needs for minerals. Attention is also focused on the significance of factors which bring about changes in the character of demand for these nonrenewable resources. The aim is to ensure the most efficient possible use of minerals.

Classification of minerals. Minerals may be conveniently classified into conventional fuels, nuclear fuels, nonfuel metallics, and other nonfuel minerals. F. Blondel and E. Ventura estimated that in 1953 the value of the world's mineral production was $36,500,000,000. Of this total, fuels accounted for 73.9%, metallic ores (excluding radioactive ores) for 21.5%, and nonmetallic for the remaining 4.6%.

Following are the 10 leading minerals by value of production in 1953 together with their proportion of the value of the world's mineral output: crud[illegible] anthracite ([illegible] gold [illegible] fuels 1.69%; lignite 1.66%; lead 1.28%; and [illegible] 1.08%. Although there are a large number of minerals, each with a variety of uses, many minerals have some uses that are similar (Table 1).

Mineral resources problems. Various problems are created because many uses are made of a single mineral, or because there is competition between minerals for a particular use. Also, minerals are nonrenewable, "fund" resources and thus for practical purposes their supply is fixed. Admittedly man can create synthetic diamonds, sapphires, mica and other minerals, but only at high economic and energy costs.

Petroleum is an example of an exhaustible fund resource, for the supply may be exhausted through use. Rain water is a typical flow resource which is renewed naturally despite its use by man. All fund resources are not exhausted through use, and those available for reuse are called revolving fund resources. The same minerals may be classed as exhaustible fund resources in certain uses, for example, lead in gasoline, and as revolving fund resources in other uses, for example, lead in car batteries. Problems are particularly acute where exhaustible fund resources are in great demand but in short supply. In such cases attempts are made to discover and use lower-priced substitutes which are in greater supply.

There is inadequate knowledge of mineral deposits which can be developed economically with present technologies. Moreover, there is much less information about mineral reserves which await technological advances and production cost reductions.

The United States has one of the world's largest and most diverse stores of minerals. Nevertheless

Table 1. General classification of selected minerals and some of their uses

Minerals	Uses
Conventional fuels	
Bituminous and anthracite coal	Direct fuel, thermal electricity, gas, chemicals
Lignite	Electricity, gas, chemicals
Petroleum	Gasoline, other fuels, chemicals, plastics
Natural gas	Fuel
Nuclear fuels	
Uranium	Nuclear bombs, power, tinting glass
Thorium	Nuclear bombs, power, gas mantles
Metallic minerals (nonfuel)	
Ferrous	
Iron	Steel
Ferroalloys	
Manganese	Alloy steels, disinfectants
Cobalt	Alloys, catalysts, radiographic, therapeutic
Columbium (niobium)	Stainless steels, nuclear reactors
Chromium	Metallurgical, refractory, and chemical
Molybdenum	Alloy steels
Nickel	In over 3000 alloys
Tungsten	Alloys and chemicals
Vanadium	Alloys
Nonferrous metals	
Copper	Electrical products, alloys
Lead	Batteries, gasoline, paints, alloys
Tin	Tinplate, solder, chemicals
Zinc	Galvanizing, die casting, chemicals
Light metals	
Aluminum	Aircraft, rockets, building materials, liquid fuel
Magnesium	Structural, refractories
Titanium	Pigments, aircraft, alloys
Zirconium	Refractories, ceramics, metals, chemicals
Other metals	
Beryllium	Copper alloys, refractories, atomic energy field
Gold	Monetary, jewelry, dentistry
Radium	Medical, industrial radiography
Nonmetallic minerals	
Asbestos	Insulation, textiles
Borides	Glass, ceramics, gasoline, solid propellants
Corundum	Abrasives
Feldspar	Ceramic flux, artificial teeth
Fluorspar	Flux, acid, refrigerants, propellants
Phosphates	Fertilizer, chemicals
Salt	Chemicals, glass, metallurgical
Sulfur	Fertilizer, acid, iron and steel industries

the supply of many essential minerals is inadequate (Table 2). Improved, more economical methods of extracting desirable minerals from low-grade ores (beneficiation) are being intensively investigated (see ORE DRESSING).

The search for minerals. Exploration for minerals could be given better direction through the development of a comprehensive program of integrated government and private research. Data should be collected and analyzed on reserves, costs and rates of exploration, and the development of mineral deposits, as well as on growing or po- new uses for known minerals. Research may also reveal new minerals of benefit to mankind.

The search for minerals can be facilitated by the development of more scientific exploration methods. Topographic and geologic mapping need to be extended and refined (see TOPOGRAPHIC SURVEYING AND MAPPING). Already geochemical and geophysical techniques have proved their value in revealing reserves of petroleum (see GEOCHEMICAL PROSPECTING; GEOPHYSICAL EXPLORATION).

Further stimulus to exploration may be provided by the free market, in which attractive prices tend to be paid for desired minerals in short supply. Increased exploration for selected minerals may result from government incentives, such as tax rebates, bonuses, loans, price stabilization, and premium price plans.

Table 2. United States supply position of selected mineral materials, 1952*

Known economic reserves adequate for well over 25 years
Magnesium, molybdenum, coal, phosphate, potash, lime, salt, sand, clay, gypsum, borax, barite, feldspar
Known economic reserves inadequate:
Discoveries geologically likely:
Copper, lead zinc, uranium, vanadium, tungsten, antimony, petroleum, natural gas, sulfur
Beneficiation progress expected
Iron, aluminum, titanium, beryllium, thorium, oil from shale, fluorine, graphite
Synthesis progress expected:
Oil from coal, gas from coal
Little or no known economic reserves, significant discoveries not expected.
Beneficiation progress expected:
Manganese
Synthesis progress expected:
Industrial diamonds, sheet mica, quartz crystals, asbestos
Significant beneficiation or synthesis not expected.
Chromium, nickel, tin, cobalt, platinum, mercury

*U S President's Materials Policy Commission, *Resources for Freedom*, vol 1, 1952

Many economists believe that future supplies of minerals will be strongly influenced by what can be paid for them, not only in money but also in human effort, capital outlay and other productive energies. Economists also believe that real costs tend to rise after the best and most accessible mineral deposits are used up. To reverse or slow this movement, technological innovations are needed to lower exploration and mining costs. Greater use must be made of lower cost substitutes (sand, clay, stone) and of foreign minerals.

Indispensability of minerals. Minerals of various kinds are indispensable for modern industrial societies to a degree dependent on the level of technology, number of people, per capita purchasing power, and consumption habits.

The supply of minerals required to satisfy the demand of each society is complicated by the uneven but fixed geographical distribution of minerals

over the earth; by the uneven geographical distribution of technical skills, enterprise, and capital necessary for the mining and utilization of minerals; and by the patterns of political boundaries. To varying degrees, the distributions of factors in the second consideration are alterable, as are political boundaries. As a result of the combination of these factors, no one country today is self-sufficient in minerals.

Demand for minerals. Countries tend to consume more minerals per capita as their population grows and they advance industrially. For example, the United States doubled its population between 1900 and 1950, but its consumption of minerals rose six times the 1900 totals in the same period. World population in 1900 and 1950 was estimated to be respectively 1,550,000,000 and 2,454,000,000, and if present growth continues, there will be at least 5,000,000,000 people in the world in the year 2000. It is likely that most countries in the future will try to raise their standard of living at a faster rate than in the past. Even if this is not possible, population growth alone will no doubt result in greater and greater demand for minerals.

Total demand for minerals in the United States will probably double between 1950 and 1975. However, demand for many essential minerals is expected to grow more rapidly in the rest of the Free World than in the United States (Table 3).

Factors influencing the accuracy of these demand predictions include changing price relationships, expanding technologies, new uses for known minerals, and the discovery of new minerals. Increases in the real costs of producing mineral supplies or substitutes must be kept below population growth; the supplies themselves must at least keep pace with the increase in population. If they do not, living standards will be undermined and economic expansion impaired, resulting in a weakening of the economic basis of national security.

Mineral conservation measures. Measures can be taken to see that the most efficient and economical use is made of minerals and that the supply of domestic and foreign minerals is improved. Technology plays a leading role in the accomplishment of such objectives. For example, technological advances since the early 1940s have resulted in the development of new ways of drilling blast holes in the troublesome, low-grade taconites (hard, iron-containing rock). Progress in flotation and magnetic beneficiation processes has helped to provide growing amounts of taconite and jasper concentrates from the Lake Superior region. The United States has thus added appreciably to its supplies of usable iron ore. Technology also helps in the development and use of renewable resource substitutes for scarce minerals and in the substitution of minerals which are more abundant. In addition, technology helps in the reconcentration of used minerals, thus permitting waste reduction and a wider application of the principle of recycling. Greater amounts of minerals can be recovered from waste products and effluents, either directly or as by-products.

The nonrenewable nature of mineral resources and the rapid depletion of so many of them necessitate reduction of the tremendous existing wastage in their production and use, a reduction that technological advances will help ensure. Legislation controlling various aspects of production may also increase usable supplies by reducing physical waste. Many states have passed laws to prevent oil wells from being spaced too close and pumped too fast, as the result is a considerable loss of reservoir energy or gas drive. Restrictions on the gas-oil ratio permitted have led to the return of excess gas to the reservoir. Such regulations, as well as technological advances such as gas and water injection to maintain pressure, have aided recovery of greater proportions of the oil in reservoirs. Other ways to enlarge mineral supplies for the future include improvement of scrap collection and changes in consumer preferences away from items using scarce minerals.

Minerals and world affairs. Problems in world affairs result from the indispensability of minerals for economic growth and their uneven geographical distribution. International trade in minerals is essential if economic growth is to continue throughout the world. Barriers to international trade in minerals, such as tariffs, export and import quotas, currency restrictions, and cartel agreements, add to problems of balancing the world's demands for and supplies of minerals.

The United States, as the world's largest single importer of minerals, has a great responsibility to stimulate the freer flow of minerals. It would also pay the United States to reduce the barriers to expansion of world trade, which would encourage the buying of minerals from the lowest cost sources and enable market forces to adjust supply to demand.

However, there are other economic and strategic aspects to be evaluated when considering measures to free world trade in minerals. Complete free trade, bringing successful competition by lower cost foreign producers, could cause the collapse of

Table 3. Estimated increase in demand, for selected minerals, 1950–1975

Mineral	United States, %	Rest of Free World, %
Aluminum	291	415
Antimony	81	100
Chromite	100	100
Coal	56	?
Cobalt	340	340
Copper	43	54
Fluorspar	187	260
Iron ore (50% Fe)	54	73
Lead	53	78
Manganese ore (46% Mn)	50	65
Molybdenum	170	170
Nickel	100	100
Petroleum	109	275
Sulfur	110	110
Tin	18	50
Tungsten	150	150
Zinc	39	61

arked diminution of the domestic minerals industry, leaving the domestic market open to exploitation by foreign producers. In addition, unemployment of local, if not national, significance could result. Domestic production of strategic minerals might be reduced to such an extent that it could not expand sufficiently in time of war. Furthermore, in wartime supply lines to foreign sources are particularly vulnerable. One of the counter arguments, however, is that domestic supplies should be conserved as much as possible in case war should break out.

The United States is particularly concerned about the short domestic supply of many strategic minerals shown in Table 2. This country is increasingly dependent on imports. For example, oil imports made up 8% of total domestic demand in 1947; in 1956 they accounted for 16.3%. In 1950 foreign iron ore made up 8% of the total used; it is expected that in 1975 foreign ore will account for 37% of the total used.

The United States and other countries attempt to ensure adequate supplies of strategic minerals by carefully balancing combinations of measures, including the promotion of accelerated growth of domestic supplies, stockpiling, and setting up stand-by facilities and reserves of minerals, as well as by developing new technologies to improve production, processing, and uses.

Demand and supply problems and the conservationist. The search for solutions to problems which arise from imbalances between mineral supply and demand is of particular concern to the conservationist. He invariably uses the ecological approach, thus giving due cognizance to the interrelationships and interdependences existing between various minerals and between minerals and other resources. Solutions for the problems will differ under conditions of peace, cold war, and hot war. Other factors influencing the formulation and character of solutions include the economic level and the rate and direction of economic growth of the economies concerned.

The cost of utilizing minerals needs further study. This includes the need to reduce further the destruction of land by open-pit mining (*see* MINING, OPEN-CUT OR PIT); the prevention of pollution of rivers and air so as to reduce the harmful effects of such pollution on health, scenery, recreation facilities; and the reuse of water downstream (*see* WATER POLLUTION).

Some people feel that technology will provide all the solutions to problems which spring from increasing demands for nonrenewable resources of limited supply. Others are much less optimistic. See separate articles for other important aspects of conservation. [M.D.T.]

Bibliography: *See* CONSERVATION OF RESOURCES.

Mineral spring

A spring, with water containing dissolved solids or gases, or which has a temperature and other physical characteristics noticeably different from the ordinary ground waters of the region in which it occurs (*see* THERMAL SPRING). Probably the best known mineral springs are those that have been used since the dawn of history in the treatment of disease or crippling ailments. Many mineral springs are remarkable chiefly for the deposits of travertine that have accumulated in fantastic forms around their orifices.

No absolute line of demarcation can be drawn between ordinary waters and mineral waters. Most mineral springs have more mineral content than springs that are used for ordinary drinking water, but some have less, and many are notable for the small amount of dissolved matter. They may be either cold or hot. Most carbonated waters are considered mineral waters, as are most sulfureted waters, even though their mineral content may be lower than that of normal springs. All natural water contains dissolved mineral matter in some degree. The kind of mineral matter and also its abundance depend chiefly upon the character of the rock through which it has passed and upon the time that elapses between its entry into the earth and its discharge. To a small degree it depends upon the amount and character of the material that is added from magmatic sources. As in normal water, the most common chemical constituents are silica, calcium, magnesium, sodium, potassium, bicarbonate, sulfate, and chloride. Iron, manganese, carbonate and many other minor constituents are present in usually small, but occasionally large quantities.

Mineral waters, used under the direction of a competent physician at mineral springs and health resorts, are generally considered to have beneficial effects. Especially in Europe such resorts constitute a large and prosperous business. In the United States the therapeutic value of mineral springs is less extensively utilized now than at the turn of the century, but several places have a large clientele, for example, Saratoga Springs, New York; Warm Springs, Georgia; and Hot Springs, Virginia. The use of mineral waters is often accompanied by marked physiological effects and when properly used such waters may be highly beneficial. When they are improperly used the effect may be deleterious. [A.N.S.]

Mineralogy

The systematic study of minerals—the materials of which the rocks of the earth's crust are made. *See* MINERAL. The study of mineralogy is limited to materials of natural occurrence but excludes the study of substances resulting from the processes of plant and animal life, such as coal, oil, amber, pearl, and bones of animals.

Mineralogy may be divided into such fields as crystallography, physical mineralogy, optical mineralogy, and chemical mineralogy.

Crystallography. With few exceptions, minerals are solids, and in most cases are crystalline, with their constituent atoms arranged in a precise geometric pattern. Because of this internal arrange-

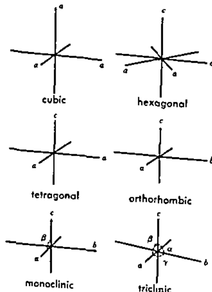


Fig 1. Crystallographic axes for the six crystal systems.

ment, minerals frequently occur as crystals. A crystal is a solid bounded by natural plane surfaces, called faces A crystal may be so characteristic in appearance that this feature alone is sufficient for identification. Moreover, many physical properties used for identifying minerals are dependent upon the crystal structure. For this reason, crystallography has largely developed as a branch of mineralogy. *See* CRYSTALLOGRAPHY.

Crystal systems. Crystals are divided into six systems (Fig. 1), on the basis of imaginary axes of reference, as follows:

Cubic: three equal axes, all perpendicular; a, a, a.

Hexagonal: three equal axes at 60°, with a fourth perpendicular to these three; a, a, a, c.

Tetragonal: two equal axes at 90°, with a third perpendicular to these two; a, a, c.

Orthorhombic: three unequal axes, all perpendicular; a, b, c.

Monoclinic: three unequal axes; one perpendicular to two inclined to each other; a, b, c, β.

Triclinic: three unequal axes, three unequal angles; a, b, c, α, β, γ.

In the hexagonal and tetragonal systems the axial ratio is c/a, and may be greater or less than unity. In the last three systems the axial ratio is $a:b:c$, with $b = 1$. In the monoclinic the obtuse angle is β; and the three triclinic angles are $b \wedge c = \alpha$, $a \wedge c = \beta$, and $a \wedge b = \gamma$.

Classes of symmetry. The six systems are divided into classes on the basis of symmetry. There are three [illegible] ter. A [illegible] half [illegible] (Fig. 2). Symmetry axes are axes about which a crystal may be rotated either 60°, 90°, 120° or 180°, and then occupy a position identical with the starting position. Such symmetry axes are known as 6-, 4-, 3-, or 2-fold axes. A crystal has a center of symmetry if for every face, edge, or corner there is a corresponding face, edge, or corner on the opposite side, at an equal distance from the symmetry center There are 32 symmetry classes. Each system has a class with maximum symmetry, and additional ones with lower symmetry.

Crystal faces. The axial ratios c/a and $a:b:c$ may have any values, and are irrational numbers. However, crystal faces always have rational intercepts in terms of the axial unit lengths, and are defined by these intercepts. The faces are designated by their parametral ratios (direct ratio of intercepts), or by Miller indices (reciprocals of intercepts, cleared of fractions). For example, an orthorhombic substance might show various crystal faces, with the parametral ratios of $a:b:c$, $a:b:\frac{1}{2}c$, $a:b:\propto c$, and $a:\propto b:\propto c$. The reciprocals of these parameters, cleared of fractions, are 111, 112, 110 and 100. As seen in Fig. 3*a*, in addition to the face labeled 111, there are seven other similar faces. These eight faces constitute a form, designated by the symbol {111}, and called the unit bipyramid. Figure 3*b* shows {111} combined with {112}, a modified bipyramid. Figure 3*c* represents {111} and the prism {110}, with four faces, while Fig. 3*d* shows {111} and {100}, the front pinacoid, with only two faces.

In an ideal crystal all faces of a given form will have the same size and shape. Crystal growth usually takes place more rapidly in some directions than in others, resulting in distorted crystals. This distortion is only in the size and shape, and never in the angular position. The angles remain constant, for the faces are parallel to internal atomic planes. *See* CRYSTAL GROWTH.

Twin crystals. These are compound crystals consisting of (1) two or more parts which seem to have interpenetrated (penetration twin) and (2) two parts, one of which seems to have been rotated 180° with respect to the other (contact

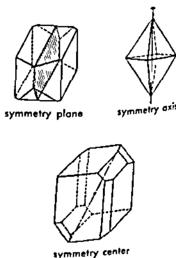


Fig. 2. Kinds of crystal symmetry. (From C. S. Hurlbut Jr., Dana's Manual of Mineralogy, 16th ed., Wiley 1952)

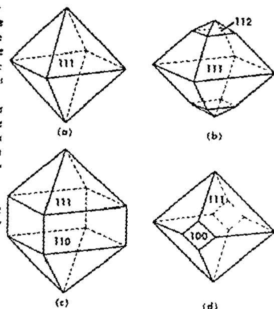


Fig. 3. Examples of orthorhombic crystal forms. (a) Unit bipyramid {111}. (b) Unit bipyramid and modified bipyramid {112}. (c) Unit bipyramid and prism {110}. (d) Unit bipyramid and front pinacoid (100)

twin), as illustrated in Fig. 4. Twin crystals are common in some minerals, and in many cases are very characteristic. *See* FELDSPAR; FLUORITE, PYRITE; SPINEL (MINERAL); TWINNING (CRYSTALLOGRAPHY).

Physical mineralogy. The major physical properties useful in mineral identification are here described.

Hardness. This important property in the recognition of minerals is defined as resistance to scratching or abrasion. A harder mineral will scratch a softer, and a softer will rub off on a harder one. Hardness is described by the numbers 1 to 10, corresponding to the hardness of ten minerals known as Mohs scale, which consists of (1) talc, (2) gypsum, (3) calcite, (4) fluorite, (5) apatite, (6) feldspar, (7) quartz, (8) topaz, (9) corundum, and (10) diamond, arranged in order of increasing hardness. Two minerals with the same hardness will not scratch each other. For example, if a mineral will scratch No 3, but is scratched by No. 4, it is said to have a hardness of 3½. Approximate hardness may be determined by using the finger nail (2½), a penny (3) or a knife blade (5½–6).

Cleavage. The splitting of a mineral in some crystallographic direction is termed cleavage. It is named according to the crystal face to which it is parallel. Not all minerals show cleavage. The quality of cleavage is described by such terms as perfect, good, indistinct, and imperfect

Fracture. Separation other than cleavage is referred to as fracture. It is described by the nature of the surface, in such terms as even, splintery, hackly, conchoidal (shell-like), rough, and smooth.

Luster. The luster of a mineral is the appearance of its surface in reflected light. The two major divisions are metallic and nonmetallic. The first refers to metals like copper as well as to such minerals as galena and pyrite which look like metals Some minerals are submetallic, and all others are nonmetallic in luster. Nonmetallic lusters are described as vitreous (glass or quartz), resinous (sphalerite), or adamantine (diamond). Fibrous minerals may have a silky luster; wavy fibers have a chatoyant luster, and a layer structure may produce a pearly luster.

Color. Some minerals have a constant color, and are termed idiochromatic. Thus the copper minerals azurite and malachite are always blue and green, respectively. Minerals with variable colors are termed allochromatic. Quartz is a striking example, coming in a wide range of colors. Internal flashes of color, as in precious opal, are called play of colors, when somewhat milky this is called opalescence.

Diaphaneity refers to transparency. A mineral description usually includes the appropriate designation as opaque, translucent, or transparent. *See* GEM.

Streak. The color of the fine powder of a mineral is known as its streak, and it may be different from that of the original mass. The streak is obtained by rubbing the specimen on unglazed white

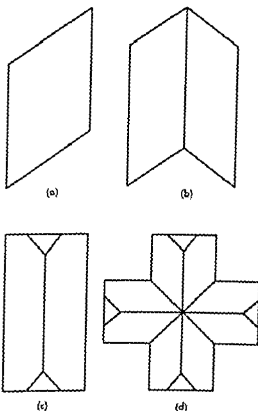


Fig. 4. Examples of twinning. (a) Gypsum single crystal. (b) Gypsum contact twin (c) Staurolite single crystal. (d) Staurolite penetration twin.

porcelain, on which it leaves a streak of fine particles. Minerals harder than the streak plate are described as having a colorless streak.

Specific gravity. Specific gravity is the ratio of the weight of a substance to the weight of an equal volume of water (at 4°C). It is obtained by weighing a mineral in air, and dividing this value by the loss of weight when weighed in water. This loss of weight is equal to the weight of the water displaced and hence equal to the volume. If a mineral is soluble in water, other liquids may be used with appropriate corrections The Jolly and the Berman balances are commonly used, as well as a chemical balance and a pycnometer bottle. *See* DENSITY MEASUREMENT; SPECIFIC GRAVITY.

Magnetism. Magnetite, pyrrhotite, and a few other minerals, when brought close to a horseshoe magnet swinging freely from a finger, will deflect the magnet Powerful electromagnets are used both for commercial separation of ores and for purifying samples in the laboratory. *See* MAGNETIC SEPARATION METHODS; ORE DRESSING; PETROGRAPHY.

Electrical properties Both electrical conductivity, as in copper, and insulating ability, as in mica, are of great practical importance, but of *no value in mineral identification.* Substances with crystal structures which have polar axes (opposite ends different) may be pyro- or piezoelectric. This refers to the development of + and − charges at opposite ends of the axis when the crystal is either heated or placed under pressure. While of no value in the identification of minerals, piezoelectricity is of great importance for frequency control. *See* PIEZOELECTRICITY; PYROELECTRICITY; QUARTZ.

Structure. Minerals may be divided into two groups: (1) amorphous, devoid of internal atomic arrangement, and hence never showing crystal faces, cleavage, or other features dependent upon structure; (2) crystalline, having an internal atomic arrangement, hence possibly showing crystal faces, cleavage, and other evidences of an atomic arrangement.

The general term structure is applied to various macroscopic features visible on a mineral specimen. The descriptive terms used are mostly self-explanatory, as fine or coarse granular, foliated, platy, scaly, fibrous, columnar, earthy, and cellular. Specimens made up of tiny spherical particles are oolitic (size of fish eggs), pisolitic or botryoidal (size of peas or grapes). Larger curved surfaces are termed reniform or mamillary, and frequently show a radial fibrous structure. Other terms used are acicular (needles), dendritic (branching), and nodular (large irregular rounded masses). Substances which are crystalline, but so fine-grained that there is no evidence of crystallinity, are [illegible] fragments [illegible] udied with [illegible] a rotating [illegible] There are [illegible], the polarizer beneath the stage, and the analyzer above, with their vibration directions at right angles to each other. In ordinary light, characteristic shapes, cleavage cracks, colors, and other details may be seen, even on microscopic grains. For measuring indices of refraction, only the polarizer is used. The index for light vibrating in one direction in the substance is thus obtained, while the rotating stage makes it possible to change directions. Cubic and amorphous substances have one index of refraction, and are singly refracting; all others are doubly refracting. Hexagonal and tetragonal substances have two, and the remaining systems have three indices of refraction. These are measured by immersing the grains in liquids of varying known index until a match is obtained. The indices of refraction are usually listed as part of a mineral description, and by themselves may be sufficient for identification. By using the analyzer as well (crossed nicols) other effects such as interference colors and figures are obtained. The latter may be uniaxial (from hexagonal and tetragonal substances) or biaxial (from the remaining systems). In twin crystals, the twinned parts have different optical orientations, and may easily be studied by optical methods. See CRYSTAL OPTICS.

Chemical mineralogy. Minerals are either elements or chemical compounds, and their formulas are determined by standard procedures of quantitative analysis. Below are given analyses of the minerals stibnite (antimony trisulfide) and calcite (calcium carbonate), showing the relative percentages found, and the manner in which formulas are derived.

Stibnite, Sb_2S_3

	(Analysis, wt %)	+ (Atomic weight)	= (Combining proportions)	Ratio
Sb	71.51	121.76	0.588	2
S	28.66	32.06	0.891	3
Total	100.17			

Calcite, $(Ca,Mg)CO_3$

	(Analysis, wt %)	+ (Atomic weight)	= (Combining proportions)		Ratio
CaO	55.38	56.08	0.988	1.005	1
MgO	0.58	40.32	0.017		1
CO_2	43.69	44.01	0.993		
Total	99.65				

In the case of calcite, the Mg is not regarded as an impurity but as Mg replacing Ca (magnesian calcite).

Chemical composition and crystal structure In many cases minerals with analogous compositions crystallize in very similar forms. This is called isomorphism. Thus the calcite group consists of calcium carbonate, $CaCO_3$; magnesium carbonate, $MgCO_3$; manganese carbonate, $MnCO_3$; iron(II) carbonate, $FeCO_3$; and zinc carbonate, $ZnCO_3$. These all crystallize in the hexagonal system, with similar axial ratios, crystal forms, cleavage, and optical properties. In such isomorphous groups

whenever the ionic radii are very close, it is possible for one ion to proxy for another in the structure, giving rise to solid solution or isomorphous replacement. Fe and Mn are so much alike that they form a complete series from $FeCO_3$ through $(Fe,Mn)CO_3$ and $(Mn,Fe)CO_3$ to $MnCO_3$. In this same group Ca and Mg are so different in size that only very limited replacement occurs. However, they do form a definite compound, $CaMg(CO_3)_2$, dolomite, in which the Ca and Mg ions alternate in regular order. The garnet group is an excellent example of isomorphism. The general formula is $M_3^{2+}M_2^{3+}(SiO_4)_3$, in which M^{2+} may be Ca, Mg, Fe, or Mn, while M^{3+} may be Fe, Al, or Cr. Isomorphism may also occur with similar anions, as PO_4 and AsO_4.

Polymorphism is the occurrence of one substance in two or more crystal structures. Usually only one of these is stable under normal conditions of temperature and pressure. In some cases two or more forms may both occur in nature, as the pairs diamond and graphite, C; pyrite and marcasite, FeS_2; and calcite and aragonite, $CaCO_3$. The minerals rutile, anatase, and brookite are trimorphous, all having the formula TiO_2, titanium dioxide.

Isostructural substances are those which although chemically quite different, have similar structures because of similar ionic sizes. Thus halite, NaCl, and galena, PbS, have similar structures, and both crystallize commonly in cubes and have cubical cleavage. Calcite, $CaCO_3$, and soda niter, $NaNO_3$, are a similar pair, with rhombohedral structures. These pairs are not isomorphous, and there is no solid solution between the monovalent and divalent ions. See ISOMORPHISM (CRYSTALLOGRAPHY); POLYMORPHISM (CRYSTALLOGRAPHY).

Blowpipe analysis. Various tests which require a minimum of apparatus and reagents have been developed for use in the field. The tip of a blowpipe is inserted into a small flame. When air is blown through it, a very hot flame can be deflected so as to impinge upon the sample being tested. By changing the position of the blowpipe, either oxidizing or reducing flames may be produced. A simple test is to determine the ease of melting, as compared with the scale of fusibility: (1) stibnite, 525°C; (2) chalcopyrite, 800°C; (3) almandite, 1050°C; (4) actinolite, 1200°C; (5) orthoclase, 1300°C; and (6) bronzite, 1400°C. For bronzite, only the tips of sharp splinters can be slightly rounded.

Volatile compounds of some elements give characteristic colors when the sample is placed in the blowpipe flame on a platinum, Pt, wire. Examples are the intense yellow of sodium, Na, compounds, and the red colorations of calcium, Ca; strontium Sr; and lithium, Li. Some copper compounds give a green flame. Wetting the powder with hydrochloric acid, HCl, may increase the volatility. To differentiate mixtures of flame colors, screens are available which cut out certain wavelengths.

Molten beads of borax or salt of phosphorus $(HNaNH_4PO_4 \cdot 4H_2O)$ are made in a loop of Pt wire. This bead is touched to a powdered mineral, and then heated in the blowpipe flame. Many metal oxides dissolve in the bead, and produce characteristic colors. In some cases the colors are different in oxidizing and reducing flames.

Heating a powdered sample in a slightly bent and inclined glass tube allows access of air, and hence produces oxidizing conditions, while heating in a closed tube is nonoxidizing. Cases with characteristic odors may be given off, or sublimates may collect in the upper portion of the tube. Likewise water vapor from hydrated minerals will condense. Some minerals change color when thus heated, some decrepitate, and Fe minerals may become magnetic. Detection of NO_3, F, Cl, Br, I, SO_4, and S in compounds is made possible by heating in a closed tube with $KHSO_4$, which liberates NO_2, HF, Cl, Br, I, SO_2, and H_2S, respectively.

Sublimates may be obtained on either plaster of paris or charcoal supports. A small depression is made in one end for the powdered sample. It may be heated by itself in the blowpipe flame, or be mixed with some reagent. A volatile sublimate may form on the cooler portion of the support. This may be a metal oxide, but when bismuth flux (containing I) is mixed with the sample, various colored iodides are formed as sublimates. The charcoal support aids in obtaining reducing conditions, in which case metallic globules may be obtained as well as sublimates.

Synthesis of minerals. Many minerals have been synthesized in the laboratory, in order to study their compositions and conditions of formation. This has been specially important in the crystallization of silicate minerals from melts, in order to understand better the crystallization of minerals from the magma. See SILICATE PHASE EQUILIBRIA.

Among the gem minerals, synthetic rubies, sapphires, spinel, emeralds, and rutile are produced commercially. Because quartz is of importance in electronics, methods have been developed to grow quartz crystals, in case natural material is unavailable. Similar experiments with synthetic mica have have not been so successful. Synthetic corundum (Al_2O_3) and moissanite (SiC) are produced on a large scale for abrasives and refractories. In the electrolytic production of aluminum, synthetic cryolite may be used in the bath. Electrodes of synthetic graphite are used because insufficient natural graphite is available. Synthetic diamond is produced in tiny crystals which have certain advantages over crushed natural diamond for abrasive purposes. See GEM, MANUFACTURED. [L.S.R.]

Bibliography: M. J. Buerger, *Elementary Crystallography: An Introduction to the Fundamental Geometrical Features of Crystals*, 1956; C. S. Hurlbut, Jr., *Dana's Manual of Mineralogy*, 16th ed., 1952; E. H. Kraus, W. F. Hunt, and L. S. Ramsdell, *Mineralogy: An Introduction to the Study of Minerals and Crystals*, 5th ed., 1959; C. Palache, H. Berman, and C. Frondel, *Dana's System of Mineralogy*, 2 vols., 7th ed., 1944–1951; H. Strunz, *Mineralogische Tabellen*, 3d ed., 1957.

porcelain, on which it leaves a streak of fine particles. Minerals harder than the streak plate are described as having a colorless streak.

Specific gravity. Specific gravity is the ratio of the weight of a substance to the weight of an equal volume of water (at 4°C). It is obtained by weighing a mineral in air, and dividing this value by the loss of weight when weighed in water. This loss of weight is equal to the weight of the water displaced and hence equal to the volume. If a mineral is soluble in water, other liquids may be used with appropriate corrections. The Jolly and the Berman balances are commonly used, as well as a chemical balance and a pycnometer bottle. *See* DENSITY MEASUREMENT, SPECIFIC GRAVITY.

Magnetism. Magnetite, pyrrhotite, and a few other minerals, when brought close to a horseshoe magnet swinging freely from a finger, will deflect the magnet. Powerful electromagnets are used both for commercial separation of ores and for purifying samples in the laboratory. *See* MAGNETIC SEPARATION METHODS; ORE DRESSING; PETROGRAPHY.

Electrical properties. Both electrical conductivity, as in copper, and insulating ability, as in mica, are of great practical importance, but of no value in mineral identification. Substances with crystal structures which have polar axes (opposite ends different) may be pyro- or piezoelectric. This refers to the development of + and − charges at opposite ends of the axis when the crystal is either heated or placed under pressure. While of no value in the identification of minerals, piezoelectricity is of great importance for frequency control. *See* PIEZOELECTRICITY; PYROELECTRICITY; QUARTZ.

Structure. Minerals may be divided into two groups: (1) amorphous, devoid of internal atomic arrangement, and hence never showing crystal faces, cleavage, or other features dependent upon structure; (2) crystalline, having an internal atomic arrangement, hence possibly showing crystal faces, cleavage, and other evidences of an atomic arrangement.

The general term structure is applied to various macroscopic features visible on a mineral specimen. The descriptive terms used are mostly self-explanatory, as fine or coarse granular, foliated, platy, scaly, fibrous, columnar, earthy, and cellular. Specimens made up of tiny spherical particles are oolitic (size of fish eggs), pisolitic or botryoidal (size of peas or grapes). Larger curved surfaces are termed reniform or mamillary, and frequently show a radial fibrous structure. Other terms used are acicular (needles), dendritic (branching), and nodular (large irregular rounded masses). Substances which are crystalline, but so fine-grained that there is no evidence of crystallinity, are termed cryptocrystalline.

Optical mineralogy. Crushed mineral fragments or thin sections of minerals may be studied with the petrographic microscope, which has a rotating stage, and uses plane-polarized light. There are two polarizing prisms; the polarizer beneath the stage, and the analyzer above, with their vibration directions at right angles to each other. In ordinary light, characteristic shapes, cleavage cracks, colors, and other details may be seen, even on microscopic grains. For measuring indices of refraction, only the polarizer is used. The index for light vibrating in one direction in the substance is thus obtained, while the rotating stage makes it possible to change directions. Cubic and amorphous substances have one index of refraction, and are singly refracting; all others are doubly refracting. Hexagonal and tetragonal substances have two, and the remaining systems have three indices of refraction. These are measured by immersing the grains in liquids of varying known index until a match is obtained. The indices of refraction are usually listed as part of a mineral description and by themselves may be sufficient for identification. By using the analyzer as well (crossed nicols) other effects such as interference colors and figures are obtained. The latter may be uniaxial (from hexagonal and tetragonal substances) or biaxial (from the remaining systems). In twin crystals, the twinned parts have different optical orientations, and may easily be studied by optical methods. *See* CRYSTAL OPTICS.

Chemical mineralogy. Minerals are either elements or chemical compounds, and their formulas are determined by standard procedures of quantitative analysis. Below are given analyses of the minerals stibnite (antimony trisulfide) and calcite (calcium carbonate), showing the relative percentages found, and the manner in which formulas are derived.

Stibnite, Sb_2S_3

	(Analysis, wt %)	÷ (Atomic weight)	= (Combining proportions)	Ratio
Sb	71.51	121.76	0.588	2
S	28.66	32.06	0 891	3
Total	100.17			

Calcite, $(Ca,Mg)CO_3$

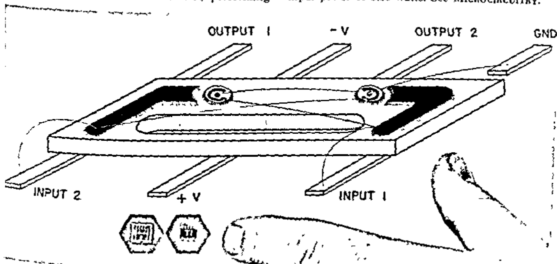
	(Analysis, wt %)	÷ (Atomic weight)	= (Combining proportions)		Ratio
CaO	55.38	56.08	0.988	1.005	1
MgO	0.58	40.32	0.017		
CO_2	43.69	44.01	0 993		1
Total	99.65				

In the case of calcite, the Mg is not regarded as an impurity but as Mg replacing *Ca* (magnesian calcite).

Chemical composition and crystal structure. In many cases minerals with analogous compositions crystallize in very similar forms. This is called isomorphism. Thus the calcite group consists of calcium carbonate, $CaCO_3$; magnesium carbonate, $MgCO_3$; manganese carbonate, $MnCO_3$; iron(II) carbonate, $FeCO_3$; and zinc carbonate, $ZnCO_3$. These all crystallize in the hexagonal system, with similar axial ratios, crystal forms, cleavage, and optical properties. In such isomorphous groups

Fig. 1. Comparison of relative size of a vacuum-tube, printed-circuit component (left), transistorized component (center), and molecular electronic component (right) satisfying similar functional specifications. (*Westinghouse Defense Products*)

an actual component assembly illustrates miniaturization in another way. A light-modulated oscillator manufactured as a vacuum-tube assembly with standard components and printed circuitry has a volume of 4 in.³, a weight of 26 g, an input power of 5 watts, 16 components, and 18 soldered connections. A transistorized version, performing the same functional role, has a volume of 1 in.³, a weight of 7 g, an input power of 0.75 watts, 14 components, and 15 soldered connections. A functionally comparable solid-state system consists of a single component with two soldered connections, a volume of 0.001 in.³, a weight of 0.02 g, and an input power of 0.06 watts. *See* MICROCIRCUITRY.



ig. 2. The semiconductor solid circuit shown above s a complete multivibrator containing two diffused-ase transistors, two capacitors, and eight resistors, all formed from a single silicon wafer. (*Texas Instruments, Inc.*)

A magnified illustration of a typical solid-state system (Fig. 2) shows a complete multivibrator circuit consisting of two diffused-base transistors, two capacitors, and eight resistors, all formed on a single silicon wafer. *See* MOLECULAR ELECTRONICS.

While electronic components provide the most dramatic examples of miniaturization, the process has been universally applied to all kinds of components and equipments found in engineering systems, including electrical, mechanical, hydraulic, and pneumatic devices.

Automated assembly. Investigations of automated assembly of electronic components have been primarily motivated by the urgent need for great quantities of electronic equipment in the event of all-out war. Some of the more highly developed and tested techniques are being applied in commercial products. Historically, many of the current concepts and techniques were tried in the 1940s. The British Sargrove machine was a large-scale attempt at automated assembly to supply the vast Chinese and Indian markets with a production potential of 500,000 radio sets a year. The process was based on a flat plastic plate on which wiring, resistors, capacitors, and inductors were formed by metal spraying and milling techniques. In the United States the first printed-circuit techniques were applied to the proximity fuse. In this case, conductive paint was deposited through a silk-screen stencil onto a ceramic plate. Commercial application of this development followed, miniature hearing-aid production being the most noteworthy example.

As mentioned earlier, the advent of automated production is influenced by, and in turn influences, the concept of dividing an equipment into subassemblies, which ultimately become relatively simple units, each a functional building block out of which component subequipments and equipments can be assembled.

Of the many automation techniques proposed and considered, two may be specially emphasized

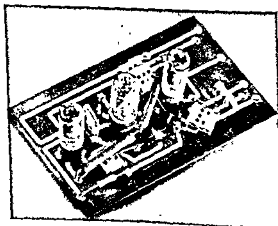


Fig. 3 Typical printed circuit with standard vacuum tubes, resistors, and capacitors. The conductors are a metallic coating on a glass base. (*Corning Glass Works*)



Fig. 4. Typical three-dimensional circuit prior to encapsulation. Unused space is reduced to a minimum by placing components in physical contact with one another. In addition, mechanical strength is enhanced by making electrical contact between elements by means of nickel ribbon which is spot-welded rather than soldered to leads of individual components. (*E. C. Hall, Instrumentation Lab, Massachusetts Institute of Technology*)

because they demonstrate two different philosophical approaches to the automated assembly problem. One approach, called Tinkertoy, broke with existing tradition and attempted to establish a new approach. The other technique, Autosembly, capitalizes upon the availability of standard components and automates the assembly of these components into multiple-element units and subassemblies.

Tinkertoy. Tinkertoy established a series of electrically and mechanically standardized modules. Each module, in general, consisted of four to six thin, 1¼ in. by 1¼ in. wafers imprinted with various circuit elements. The wafers in each module were arranged like the floors in a building. Wires laid in grooves around the periphery of the wafers and provided both vertical support for the wafers and electrical connections between various component elements on the several wafers. Each of the module assemblies culminated with a plug which was inserted into a corresponding socket on the next module; thus larger equipments could be built quite readily and flexibly. The production of the wafer assembly was completely mechanized.

Autosembly. Autosembly consists of assembling conventional components to printed conductor panels by dip soldering. An equipment might employ several such panels spaced apart like the decks of a ship with headroom for the components determining the panel spacing. Interpanel connectors provide both circuit and mechanical connections and are secured to the panels in the same manner as other components. A typical example of a printed-circuit panel with conventional components is shown in Fig. 3. *See* PRINTED CIRCUIT.

Encapsulation. Using the technique of encapsulation, an entire circuit or assembly is embedded in a casting resin. This technique is similar to the potting of transformers and filters in asphaltic materials such as tar and pitch. Encapsulation gives structural integrity to the assembly, increasing its resistance to shock and vibration, protecting it from moisture, fungus, dirt, and other foreign matter

rials, and often reducing the weight of the assembly by eliminating chassis, bolts, and other hardware. Encapsulation is used on both panel-mounted printed circuitry and dense-packed point-to-point wired equipment, as shown in Fig. 4. Plastic polymers are usually used as the encapsulating material. These substances are viscous liquids in the unpolymerized state, but when set, they harden into a rigid mass. The encapsulation process is usually conducted in a vacuum to avoid air entrainment and to facilitate quick and easy flow of the resin. In addition to satisfying castability, casting resins must have mechanical characteristics of adequate strength, light weight, and tight sealing, and electrical properties that will not cause deterioration of the performance of the assembly.

Such encapsulated components or equipments are usually unrepairable because the casting resin often cannot be removed without damage to the components and wiring. This possible disadvantage is often overcome by subdividing the equipment into subcomponent assemblies, which are discarded when defective. *See* MAINTAINABILITY OF EQUIPMENT; SYSTEMS ENGINEERING. [R.W.M.]

Minimal principles

In the treatment of physical phenomena, it can sometimes be shown that, of all the processes or conditions which might occur, the ones actually occurring are those for which some characteristic physical quantity assumes a minimum value. These processes or conditions are known as minimal principles. The application of minimal principles provides a powerful method of attacking certain problems that would prove formidable if approached directly from first principles.

One simple minimal principle asserts that the state of stable equilibrium of any mechanical system is the state for which the potential energy is a minimum. Other general theorems of classical dynamics that are related to minimal principles are Hamilton's principle and the principle of least action (*see* HAMILTON'S PRINCIPLE; LEAST ACTION, PRINCIPLE OF).

Minimal principles are important in branches of physics other than mechanics. Fermat's principle in optics, for example, states that, of all possible paths of light transmission between two points, the actual path is the path for which the transmission time is a minimum. Minimal principles also find wide application in thermodynamics. [D.WI.]

Bibliography: R. B. Lindsay, *Concepts and Methods of Theoretical Physics*, 1951.

Mining

The process of extracting mineral wealth from the earth, sea, or atmosphere. The term embraces the recovery of oil and gas from wells; metal and nonmetallic minerals from the earth and sea; sulfur, salt, and other brines from wells; and coal, peat, oil shale, tars, and other hydrocarbons from the earth. As used in this article, mines and mining embrace the entire mineral industry. For other aspects *see* COAL MINING; IRON (EXTRACTION FROM ORE); METALLURGY; ORE DRESSING.

Mineral deposits constitute one of the great classes of natural resources. Unlike forests, fisheries, falling water (for power), and fields for agriculture and grazing, mineral deposits are not self-renewing but, as natural resources, are termed "wasting assets." This concept of wastage, however, is more apparent than real, in that the exhaustion of a high-grade deposit invariably makes possible the profitable operation of some lower-grade deposit previously economically submarginal. Moreover, many minerals are subject to reclamation and may be reused with little irretrievable loss each time. Nevertheless, prudence on the part of nations and individuals calls for the conservative exploitation of all natural resources so that they may serve the highest uses of society. *See* CONSERVATION OF RESOURCES; EARTH RESOURCE PATTERNS.

The concentration of minerals in a deposit necessary to qualify it for commercial interest varies with the value of the mineral being sought. Commercial iron ore bodies are 25–65% iron; aluminum-bearing ore bodies (bauxite), 30–35% aluminum; uranium ore bodies, 0.02–2% uranium oxide; major copper ore bodies, 0.6–1.5% copper; major lead-zinc ore bodies, 3–5% metal; gold ore bodies, a few tenths of an ounce of gold per ton; placer ore bodies, as low as $0.35 worth of gold per cubic yard. Discovery of the deposits may be accidental or the result of systematic prospecting. *See* MINERAL FUEL AREAS; MINERAL RESOURCE AREAS; PROSPECTING.

The character of mineral deposits as well as the state of the technological arts influences selection of mining technique. Minerals in fluid form, or those which can be reduced to fluids by solution or heat, usually are extracted through drilled wells. Virtually all petroleum and natural gas is so extracted. Brine wells include those tapping natural brines and those injecting water and withdrawing brine from soluble deposits. Sulfur is melted underground by injection of superheated water and is pumped to the surface as a mixture of hot water and molten sulfur (*see* BORING AND DRILLING, MINERAL). Alluvium and other detrital matter containing valuable minerals are classified as placer deposits and are ordinarily mined by hand, by ma- [illegible]

than placers may be classified by type of deposit and by mining method. Bedded deposits are of great significance because they include all the coal mines and most salines, limestone, gypsum, and many others. In contrast to bedded deposits, those [illegible]

rying, from underground operations. *See* MINING, UNDERGROUND; *see also* MINING, OPEN-CUT OR PIT; QUARRYING.

Mining machinery is commonly adapted to the functions it performs. Drills are used to penetrate rock to explore its nature, to provide passageways for fluid minerals, and to permit the effective placement of explosives. Broken rock is handled with hand shovels, power scrapers and power shovels, and by gravity (*see* MINING EXCAVATION). Rock is transported by truck, train, belt, and, when major lifts are involved, by power hoist. Auxiliary mechanical equipment includes pumps, ventilators, and many other items (*see* MINING OPERATING FACILITIES).

Mining—particularly underground mining—is hazardous because of rock falls and, in many coal seams, the presence of explosive gases. Some rock dusts may cause silicosis when inhaled, and some metals including mercury and lead are poisonous if ingested or inhaled. *See* MINING SAFETY.

Differences in national laws and attitudes contribute to diversity in mining. Land tenure under mining law may reserve to the state the right to mine, as in Russia, under the doctrine that political power stems from control of means of production. At the other extreme is mining law in the United States, where public lands exclusive of certain parks, military reserves, Indian reservations, and other areas are open to prospecting. The eastern states do not come under the Federal mining law, but many have laws pertaining to the acquisition of mining rights or leases. There are two types of claims that may be staked on Federal public lands: (1) *lode claims*, measuring 600 by 1500 ft, which apply to relatively narrow vertical or sloping veins, and (2) *placer deposit claims*, measuring 20 acres. Placer deposits consist of sand, gravels, or other unconsolidated material lying on the surface.

Most national mining codes treat subsurface rights as separate from surface rights, and in most jurisdictions the sovereign claims mineral rights. Even in the United States many minerals on the public domain may only be worked privately under the Leasing Act of 1920, by which the Federal government retains ownership of certain minerals.

Control of [illegible] in the exe[illegible] mercial. N[illegible] eral supp[illegible] their bord[illegible] within their spheres of influ[illegible] [illegible] consumers recognize the trade control possibilities stemming from possession of major productive ore bodies. [C.W.M.]

Bibliography: Georgius Agricola, translated from the first Latin Edition of 1556 by H. C. Hoover and L. H. Hoover, *De re metallica*, 1950; Seeley W. Mudd (ed.), *Economics of the Mineral Industries*, 1959; *Mineral Facts and Problems*, U.S. Bureau of Mines Bulletin 556, 1956; T. A. Rickard, *Man and Metals*, 2 vols., 1932; U.S. President's Materials Policy Commission, *Resources for Freedom*: I, *Foundations for Growth and Security*, II, *The Outlook for Key Commodities*, 1952; W. Van Royen, O. Bowles, *Atlas of the World's Resources: The Mineral Resources of the World*, vol. 2, 1952.

Mining, open-cut or pit

Open-cut or pit mining is one of the outstanding methods of exploiting minerals from ore deposits that are commonly somewhat consolidated at or close to the earth's surface. This article deals mostly with metallic-ore extraction; for other near surface mining methods applied to mineral ores which are generally little consolidated, *see* MINING, PLACER; MINING, STRIP. For applications of similar methods to special problems see COAL MINING.

Most open-cut and pit mining develops an annular pattern of benches or terraces. These are cut back into and around hill or mountain ore masses or are pushed downward and sideways to cut pits into ore deposits, after any necessary removal of overburden. Overburden is waste material of unconsolidated or consolidated rock, blocking ready surface access to the ore body. Unconsolidated overburden may be readily stripped off, but all other rock and ore is excavated by varying combinations of the four principal operations of mining—drilling, blasting, loading, and haulage (*see* MINING EXCAVATION). Mine waste disposal, slope stability, and communications are special problems in relation to open-cut or pit mining.

Rock drilling for open-cut mining. Four types of machines are used to bore blast holes for explosives. The blast holes range from 2½ to 12 in. in diameter, and are drilled in the depths and directions indicated for any particular benchlike excavation pattern. *See also* BORING AND DRILLING, MINERAL.

Churn drill. Its relatively light weight and vertical drill holes from 6 to 9 in. in diameter make this type adaptable for the initial work of starting a [illegible] within an estab[illegible] [illegible]dly bench[illegible]



Fig. 1. Roads and rounds develop open cut [illegible] Bolivar in Venezuela. (U.S. Steel)



Fig. 2. Large open-pit copper mine showing rail haulage and trolley-electric locomotive. (Kennecott Copper)

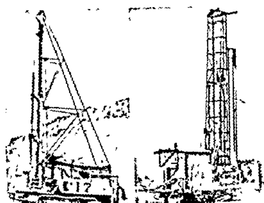


Fig. 3. Drilling machines. (a) Churn drill makes vertical blast holes. (b) Rotary drill drives vertical, large-diameter blast holes. (Kennecott Copper)

Rotary drill. Two kinds of rotary drills are used to bore vertical holes from 4½ to 12 in. in diameter. Compressed air is used to cool the bit and to carry from the hole rock cuttings which are used to refill the blast hole after the explosive charge has been placed. Smaller, truck-mounted rotary drills find popular application in initial work and in developing new levels; the larger, heavy, caterpillar-mounted machines are better for bench blasting in the pit.

Percussion drills. Where space on the pit levels and on the access roads between levels is limited, the rotary drill has a decided disadvantage, and the percussion drill becomes an important tool. Water is required to reduce dust when drilling; both water and compressed air are required to clear the bit and to wash cuttings from the hole. The machine is commonly mounted as a self-contained, mobile drilling unit. The vehicle is either a crawler or rubber-tired tractor on which all necessary equipment and supplies are mounted.

Although capable of boring either vertical, inclined, or horizontal holes, the usual application of the machine is for boring inclined holes at or near the base of a pit bench or level. Bit diameters in use range from 2½ to 6 in. Toe or inclined holes are mostly limited in practice to 35 ft in depth. These holes, because of their small diameter, must be enlarged at the bottom to provide room for the necessary amount of explosive material. This is accomplished by detonating light explosive charges at the bottom of the hole, a process known as chambering or springing.

Jet piercing. The jet piercing machine, using a jet flame of kerosine and oxygen under pressure, has been developed to penetrate vertically the very hard taconite ores. A 6½-in.-diameter hole is most common. The resulting hole is very irregular from top to bottom, which necessitates careful selection and charging of explosives. Piercing machines to drive inclined holes are in the developmental stage.

Blasting. The type and quantity of explosive are governed by the resistance of the rock to breaking. Dynamite and ammonium nitrate find wide application in open-pit mining, and the latter is becoming more common as a low-cost explosive.

Dynamite. Dynamite of varying strengths is used in open-pit mining according to rock conditions. It is common either in cartridge form or as a pulverized, free-flowing material packed in bags. In a small-diameter, chambered hole, the explosive is poured from the bag to fill the chamber.

Ammonium nitrate. Commercial, or fertilizer-grade, ammonium nitrate has recently become a popular blasting medium. It is not in itself sensitive to detonation by the conventional blasting cap and, consequently, is safer to handle, store, and transport than most explosives. The granular or prill size ammonium nitrate is commonly packed in paper, textile, or polyethylene bags. The carbon necessary for the proper detonation of ammonium nitrate is usually provided by the addition of fuel oil. Where the small-diameter, horizontal blast hole is used, the explosive is blown into the hole by compressed air. Ammonium nitrate is used extensively in blasting vertical holes, but the hole must be free of water because the material is highly soluble. A recent development is a slurry mixture of 65% ammonium nitrate, 20% TNT, and 15% water for blasting water-filled holes.

Mechanical loading. Ore- and waste-loading equipment in common use includes power shovels, for medium to large pits; and tractor, or front-end loaders, and rocker shovels for the smaller operations. The loading unit must be selected to fit

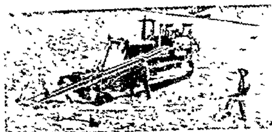


Fig. 4. Crawler-mounted drilling machine. Horizontal toe holes are drilled by percussion. (Kennecott Copper)

the transportation system. Because the rock will usually be broken to the largest size that can be handled by the crushing plant, the size and weight of the broken material indicate the proper type of loading machine. Of equal importance in determining the type of equipment is the production or loading rate. The condition of the benches or levels, and the connecting ramps and the bench room available will determine what operational mobility will be required of the machine.

Power shovels. Shovel capacities range from 2½ to 10 yd³, as boom and dipper handles also vary in length. The boom and dipper handle lengths determine the digging radius and the maximum bench height that can be safely handled.

4000 tons per 8-hour shift, while a 6-yd shovel will handle 6000–8000 tons per shovel shift. However, the nature of the broken material will affect production considerably. Power may be derived from diesel or gasoline engines, diesel-electric motors, or electric motors. The all-electric machine is most widely used in the larger pits.

Draglines. A dragline is used to move or load broken ore or waste. It consists of a long boom mounted on a mobile cab housing a hoist and power

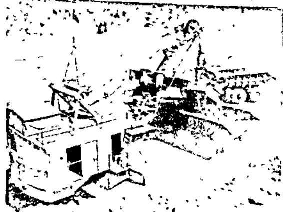


Fig 5. Loading haulage trucks in open-pit copper mine. (*Kennecott Copper*)

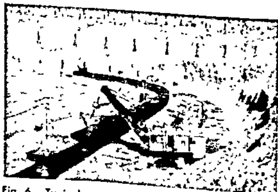


Fig 6 Typical open-pit loading. Electric shovel loads ore into railroad cars hauled by trolley-electric locomotives (*Kennecott Copper*)



Fig. 7. Small open-pit copper mine showing truck and skip-haulage systems. (*Pima Mining*)

unit. A bucket is suspended by steel cable from a sheave at the end of the boom. The bucket is cast out toward the end of the boom and pulled back by the hoist to gather up a load of ore, which is deposited in a haulage unit or on a waste pile. Draglines vary in capacity from a few yards to 30 yards.

Mechanical haulage. Rail haulage, trucks, inclined skip hoists, and belt conveyors find application under the proper conditions for transportation of ore and waste. The factors controlling the selection of a haulage system are principally the size and depth of the pit, the production required, and the length of haul to the crusher or waste dump.

Rail haulage. Where mine rock must be transported more than 1½–2 miles, rail haulage is generally employed. Since rail haulage requires a larger capital outlay for equipment than other systems, only a large ore reserve justifies the investment. As a rough rule of thumb, the reserve should be large enough to support for 25 years a production rate of 30,000 tons of ore per day and an equivalent or greater tonnage of stripping. Adverse grades should be limited to a maximum of 3% on the main lines, and 4% for short distances on switchbacks. Good track maintenance requires the use of auxiliary equipment such as mechanical tie tampers and track shifters. The latter are required for relocating track on the pit benches as the mining progresses and on the waste dumps as the disposed material builds up adjacent to the track. Ground movement in the pit resulting from disturbance of the earth's crust or settling of the waste dumps makes track maintenance a large part of mining cost.

Locomotives in use range from 50 to 125 tons in weight, with the latter coming into increased use for steeper grades and larger loads. Most mines operate either all-electric or diesel-electric models. The use of all-electric locomotives creates the problem of electrical distribution in the pit and on the waste dumps, and requires the installation of trolley lines adjacent to all tracks.

Mine cars range in capacity from 50 to 100 tons of ore, and to 40 yd³ of waste. Ore is transported

in various types of cars—solid-bottom, side-dump, or bottom-dump. The solid-bottom car is cheapest to maintain, but requires emptying by a rotary dumper. Waste is mostly handled by the side-dump car.

Truck haulage. For smaller mines and in areas of large pits where the haul length does not exceed 1½–2 miles, truck haulage is particularly applicable. Trucks may also be used for longer hauls in cases of extremely rough topography, poor mining conditions in the pit, and short mine life. They commonly supplement rail haulage and the inclined skip hoist. Adverse grades are normally limited to 8%, but most downhill grades are limited to 12%. Popular rated capacities of truck range from 22 to 55 tons. Some 70-ton haulers are employed, and the trend points toward greater use of the larger vehicles. Rear-dump and side-dump trucks are most popular. Self-loading scraper-type trucks are particularly adaptable for removing unconsolidated overburden.

Inclined skip hoist. Under certain conditions, the inclined skip hoist is replacing rail haulage in the deeper, open-pit copper mines. When the pit has reached a depth of approximately 300 ft, it is possible to substitute the inclined skip hoist for long hauls by rail or truck which spiral around the pit (see Fig. 2).

Skips range in capacity from 20 to 35 tons and are usually built to contain the load of one motor truck. The skip system is always used in conjunction with trucks or rail haulage, or both.

Conveyors. Rubber-belt conveyors may be used to transport crushed material from the pit at slope angles up to 20°. Conveyors are especially useful for transporting large tonnages over rugged terrain and out of pits where ground conditions preclude building of good haulage roads. Improved belt design is permitting greater loading, higher speeds, and the substitution of single-flight for multiple-flight installations. The chief disadvantage of this transport system is that, to protect the belt from damage by large lumps, waste as well as ore should be crushed in the pit before it is loaded on the belt.

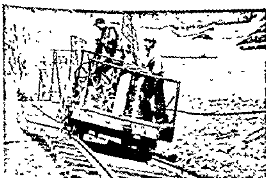


Fig. 8. Track shifter used on large rail-haulage system in open-pit mines. (Kennecott Copper)

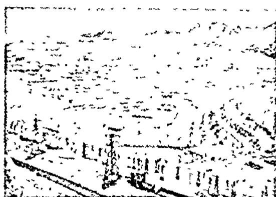


Fig. 9. Dumping copper-mine waste by trolley-electric haulage on side-dump cars. (Kennecott Copper)

Waste-disposal problems. To keep costs at a minimum, the dump site must be located as near the pit as possible. However, care must be taken to prevent location of waste dumps above possible future ore reserves. In the case of copper mines, where the waste contains quantities of the metal which can be recovered by leaching, the ground on which such waste is deposited must be impervious to leach water. Where the creation of dumps is necessary, problems of possible stream pollution, and the effect on farms and on real-estate and land values, must all be considered.

Slope stability and bench patterns. Care should be taken in designing bench cuts to provide the proper slope, that is, the angle at which the benches progress from the bottom to the top of the digging. Faults, joints, bedding planes, and especially ground water in the cut slopes are known to contribute to rock and earth slides. Mine blasting will shatter the rock and reduce its structural strength. As the excavation increases in depth, the condition of the surrounding earth mass becomes more critical to stability. In practice, overall pit slopes range from 22 to 60° according to ground conditions. Where little information is available, the pit is usually planned for an overall slope of 45°; subsequent mining experience determines the final slope.

In open-cut and pit mining the material ranges from unconsolidated surface debris to solid, thoroughly consolidated rock. Soil and rock mechanics aids in determining stable slopes. These technologies are being extended to assist in determining safe slopes in rock portions of the pit. Although it is in the experimental stage, instrumentation is being developed to detect the boundaries of moving rock masses, the rate of movement, and to forewarn of impending failure.

Communication. Efficiency of mining operations, especially loading and hauling, is being improved by the use of communication equipment. Two-way, high-frequency radiophones are proving useful for communicating with haulage and repair crews and shovel operators. [A.S.]

Mining, placer

Essentially surface mining of placer deposits of sand, gravel, and other alluvium and eluvium containing concentrates of metals or minerals of economic importance. For many years gold has been the most important placer metal mined, although large productions of platinum, tin, phosphate, monazite, columbite, ilmenite, zircon, diamond, sapphire, and other gems have come from placers. Other valuables thus derived include native bismuth, native copper, native silver, cinnabar, and other heavy weather-resistant metals or minerals. For geological and related physical character of placer deposits, *see* ORE AND MINERAL DEPOSITS.

Prospecting, sampling, and valuation. A uniform grid system pattern is best used in sampling placers. This generally consists of equally spaced sample points in lines about perpendicular to the length of the deposit. Samples are usually obtained by one or several of the following methods: shaft sinking, caisson sinking, churn drilling, and open cuts. *See* BORING AND DRILLING, MINERAL; MINING OPERATING FACILITIES; PROSPECTING.

Recovery methods. Because the proportion of gold or valuable mineral is commonly low in relation to the bulk of placer material, reduction of bulk and recovery practices are closely linked with most placer mining methods. Four variations on the flowing-water sluice-box method are outlined here; details of their design and application appear in the technical literature.

Fig. 1. Sectional diagrams of sluice construction. (*From G. J. Young, Elements of Mining, 4th ed., McGraw-Hill, 1946*)

Fig. 2. Types of riffles in partial plan views. (*From G. J. Young, Elements of Mining, 4th ed., McGraw-Hill, 1946*)

Made of 1⅛ in. stock dressed to 1 in. All joints carefully fitted and fastened with screws sunk flush. Canvas bottom stretched taut and fastened with ½ in. quarter round. Canvas bottom

Fig. 3. Plan details of a rocker. (*Modified from a drawing in Engineering and Mining Journal in R. S. Lewis, Elements of Mining, 2d ed., J. Wiley and Sons, 1941*)

Riffles. Various barriers to smooth flow, such as slats, poles, blocks, and 3–6 in. rocks, establish turbulent flow, in the quiet eddies of which heavier metals settle.

Mercury and amalgamation. Mercury, sprinkled daily at the head of sluice boxes to combine with and hold passing gold, and in shallow (¾ in., partly filled) troughs to act as amalgamating traps, is a common adjunct to sluicing. Copper plates, mercury-coated and protected from coarse gravel by suitable screen (grizzly), aid in collecting gold under some conditions.

Undercurrents. These are auxiliary sluices, parallel or at right angles to the main sluice. They use a variety of grizzly plates, screens, and mattings to separate various sizes of material and to catch small particles of concentrates that may be lost when using only a straight sluice box.

Jigs. These are essentially formed as a topless box with a screen bottom, upon which rests a 3–6 in. bed, usually of steel shot, and through which pulsating water currents act. The valuable heavy particles pass through the pulsating bed and are recovered as a concentrate. Lighter water-borne material passes over the bed and into other recovery units or emerges as a finished tailing.

Mining methods. An inclusive classification of present placer mining contains five types: small-scale hand methods, hydraulicking, mechanical equipment mining, dredging, and drift mining.

Small-scale hand methods. Perhaps most widely known because of gold-rush notoriety, the small-scale methods of panning, rockers, long tom, and a few others tend to account for a decreasingly small per cent of all placer mining in competition with those of larger and more mechanized types.

Pan and batea. Panning is the recovery of gold by use of a circular metal dish that varies in diameter from 6 to 18 in., 16 in. being quite common

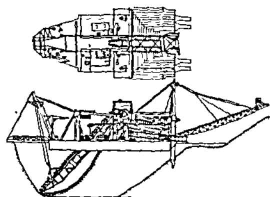


Fig 4. Plan and elevation drawings of a bucket-ladder dredge. (Yuba Manufacturing Company)

Many such pans are from 2 to 3 in. deep and have sloping sides. The pan with the mineral-bearing gravel, immersed in water, is shaken to cause the heavy material to settle toward the bottom of the pan while the light surface material is washed away by swirling and overflowing water until only the concentrate remains. At the present time, the pan is used mostly for prospecting and sampling.

In some countries a conical-shaped wood unit called a batea (12–30 in. diameter with about 150° apex angle) is used to recover valuable metals from river channels and bars.

Rocker. Rockers are used to sample placer deposits or to mine high grade areas when installation of larger equipment is not justified. Essentially, all of numerous designs consist of a screen hopper, recovery units and provisions (rocker supports) for giving the entire machine a hand rocking motion which is perpendicular to the length of the sluice portion. Placer gravel is washed through the screen with water onto a canvas apron (onto matting or mercury plates or traps, in the case of gold) and then over Hungarian riffles.

Long tom. A long tom is essentially a small sluice box with various riffles, matting, and expanded metal screens; in the case of gold, the long tom is also equipped with amalgamating plates and traps.

Other small-scale methods. These include shoveling into boxes, ground sluicing, booming, and the

Fig. 5. Aerial oblique view up-valley over placer dredging, showing successive cuts, tailings patterns, and dredge pond, dredge at left. (Pacific Aerial Survey, Inc.)

use of dip boxes, puddling boxes, dry washers, surf washers, and small-scale placer machines.

Hydraulicking. Hydraulic mining utilizes water under pressure, forced through a nozzle, to break and transport the placer gravel to the sluice box where it is washed and stacked by water under pressure. Hydraulicking is a low-cost method of mining if a cheap, plentiful supply of water is available and streams are not objectionably polluted.

Mechanical equipment. Methods under this category utilize bulldozers, draglines, pumps, and sluice boxes in a number of different combinations depending on physical characteristics of placer deposits. See BULK-HANDLING MACHINES.

Dredging. Because of ability of this method to process, fairly cheaply, huge volumes of low-mineral materials, most placer operations on large, physically favorable sands and gravels are by dredging.

Bucket-ladder dredges. This mass-handling method now largely supersedes others, such as the chain-bucket dredge, one-bucket dredge, and suction dredge for mining placer gold and other heavy placer minerals. Some large modern dredges can dig 10,000 yd³ or more per 24 hr.

Modern dredges are for the most part steel hulls of the compartment or pontoon types. On the hull is mounted the necessary machinery to cause an endless bucketline to revolve and dig the placer gravels. The buckets discharge their load into a hopper where it is washed with water into a long inclined revolving cylindrical screen having holes that commonly vary in diameter from ¼ in. to ½ in. In this revolving screen, a stationary manifold supplies water (approximately 60 psi) to nozzles spaced equally in a line the length of the screen. Water from these nozzles washes the sand and gravel and causes the fines and valuable minerals and metals to work through the holes in the revolving trommel and fall on the tables (series of parallel sluices) which may consist of various recovery units, such as mercury traps, jigs, riffles, matting, expanded metal screens, and undercurrents. The placer concentrates are held in the recovery units until clean-up time (the end of a 1 day to 4 week period). The accompanying sand passes over the tables and into the dredge pond. The coarse material that will not pass through the holes in the screen is discharged at the lower end of the trommel screen onto an endless conveyor belt which is supported by the stacker. This belt carries the coarse tailings to a final resting point in back of the dredge.

Dredges float in ponds which advance as the bucket digs the gravel while discharged sand and coarse tailings fill the pond in back of the dredge.

Dragline dredges. A washing and recovery plant, mounted on pontoons or crawler treads, is fed directly or by a conveyor from a shovel or dragline rig. Such dragline dredge installation is mostly used on deposits too small to justify the installation of a bucket dredge.

Drift Mining. Because of relatively high costs, drift mining of placer deposits is becoming less popular. It consists of sinking a shaft to bedrock, driving drifts up- and downstream in the valley deposits for 250–300 ft, and then extending crosscuts the width of the paystreak (see MINING, UNDERGROUND). In thawed ground timbering is necessary, whereas in frozen ground a minimum of timber is required and steam thawing plants are used to thaw the gravels before mining is done.

Recent improvements. In past years a number of improvements have been made to increase the efficiency of placer mining. These may be summarized as follows:

1. Development of the diesel engine allowed diesel-powered pumps, tractors, draglines, dredges, and the like to be manufactured and gave much more compactness, economy, and flexibility to placer mining methods. This in turn resulted in greater yardage at lower unit costs.

2. The adaption and use of jigs in placer recovery plants has increased the recovery of valuable minerals and metals with less labor and more efficiency of compact recovery units.

3. The use of the sluice plate in mechanical mining methods has been one of the most important improvements in recent years. This method does away with the conventional nozzle set up in front of the sluice box and the accompanying labor expense. It results in a fast system of coordinated operation, especially for mining placer gravels that contain a minimum of clay and cemented gravels.
[E.H.E.]

Bibliography: C. F. Jackson and J. B. Knaebel, *Small-Scale Placer-Mining Methods*, U.S. Bur. Mines, Inform. Circ. 6611, 1932; R. Peele and J. A. Church (eds.), *Mining Engineers' Handbook*, 3d ed., 2 vols., 1941; E. B. Wilson, *Hydraulic and Placer Mining*, 3d ed., 1918.

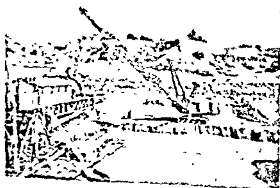


Fig. 6. Dragline dredge and pond. (Bucyrus-Erie Company)

Mining, strip

A surface method of mining by removing the material overlying the bed and loading the uncovered mineral, usually coal. It is safer than underground

ining because neither the workers nor the equip-
ient is subjected to such hazards as roof falls and xplosions caused by gas or dust ignitions. Coal ear the outcrop or at shallow depth can be tripped not only more cheaply but more completely han by deep mining, and the need for leaving illars of coal to support the mine roof is elimi-ated. The roof over coal at shallow depth is weak nd difficult to support in underground workings y conventional methods, yet this same weakness, f cover and of coal seam, makes stripping less ifficult.

Power shovels, draglines, bulldozers, and other ypes of earth-moving equipment slice a cut through he overburden down to the coal. The cut ranges rom 40 to 150 ft wide, depending on the type and ize of equipment used. The stripped overburden spoil) is stacked in a long ridge (spoil bank) arallel with the cut and as far as possible from ndisturbed overburden (highwall). The slope of a poil bank is approximately 1.4:1 and that of a ighwall under average conditions is 0.3:1. The un-overed coal (berm) is then fragmented, loaded, nd transported from the pit. Spoil from each ucceeding cut is stacked overlapping and parallel vith the previous ridge and also fills the space left y the coal removed.

Techniques of stripping methods are similar, but he size of equipment used depends on whether the nine is in prairie or hill country. In prairie areas he thickness of overburden is nearly uniform, the oal bed is extensive, and equipment can be used or years at one mine without dismantling and mov-ng to another location. Large-capacity shovels cost-ng $1,000,000 or more and requiring many months o erect on the site are used at prairie mines. A nit of this type is the 60 yd³ rig shown in Fig. 2 n 1958 the world's largest power shovel, with a lipper capacity of 70 yd³, was placed in operation t the River King mine of Peabody Coal Company near Freeburg, Illinois.

Most coal underlying hills is mined by under ground methods, but where the working adjoins the outcrop and the overburden is thin, the roof be-comes difficult and expensive to support. The coal between the actual or potential underground work-

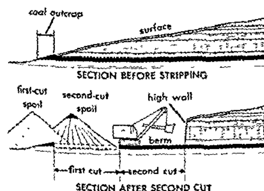


Fig 1. Representative cross-section profile diagrams of contour strip mining of coal.

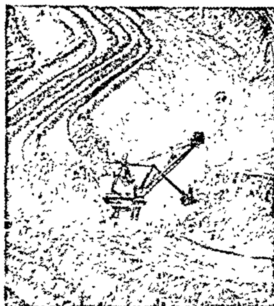


Fig 2 Large electric shovel, high wall (right), and spoil bank (left) at Hanna Coal Company's George-town mine, eastern Ohio (U S. Bureau of Mines and Marion Power Shovel Co.)

ings and the outcrop is then available for stripping. Usually only two or three cuts 40–50 ft wide can be made on the contour of the coal bed, after which the shovel has to be moved to another site. Thus, in contour stripping, mobility of shovels up to 5 yd³ capacity is more important than large capacity. Large draglines are used instead of shovels to strip pitching beds of anthracite to depths surpassing 400 ft, but the method could more properly be classed as open pit. See MINING, OPEN-CUT OR PIT.

Removal of unconsolidated overburden by hy-draulic monitoring is a technique used especially in Alaska. Water under a high-pressure head is di-rected through a nozzle against the overburden to wash it into deep valleys where swift streams carry it away.

Although the character of the overburden deter-mines the thickness of overburden that can be stripped, the maximum for shovels up to 5 yd³ ca-pacity is about 50 ft and for the largest equipment about 110 ft To reach these goals it frequently is necessary to use a dragline carryall or bulldozer on the highwall or a dragline in tandem with the shovel on the berm to strip the upper few feet of over-burden.

Rocks overlying coal beds present some diversity of conditions for removal. Materials generally com-prise shale, sandstone, and limestone with shale predominating. Proper fragmentation before strip-ping may be necessary to produce sizes that are smaller than the shovel dipper. Probably more re-search has been done on overburden drilling and blasting than on any other phase of stripping. The diameter, depth, and spacing of drill holes, the type of drill (whether vertical or horizontal), and

the amount and type of explosive for each blast hole are the variables that must be determined for optimum production. Truck-mounted rotary drills have replaced churn drills for drilling vertical holes, and for horizontal drilling, auger drills are used. Package explosives, Airdox, Cardox, and commercial ammonium nitrate mixed with diesel fuel are used for blasting. The ammonium nitrate–diesel fuel mixture is one of the cheaper explosives and has gained favor rapidly. Equipment for mixing this explosive and automatically injecting it through a plastic tube into horizontal drill holes is being tested. If perfected, this will mechanize the only manual operation remaining in the stripping cycle.

Before coal is loaded, spoil remaining on the berm is removed by bulldozer, grader and rotary brooms, or at small mines by hand brooms. Small-diameter vertical holes are augered into the bed and blasted with small charges of explosive to crack the coal. A ripper pulled by a bulldozer is effective in replacing coal drilling and blasting. Broken coal is loaded by ½ to 5 yd^3 capacity shovels into trucks of 5 to 80 ton capacity and transported from the stripping. [J.J.D.]

Bibliography: *Mining Guidebook and Buying Directory* (annual); R. Peele, *Mining Engineers' Handbook*, 3d ed., 1948; A. L. Toenges et al., *Some Aspects of Strip Mining of Bituminous Coal in Central and South Central States*, Information Circular 6959, U.S. Bureau of Mines, 1937.

Mining, underground

One of the principal methods of extracting minerals and ores from distinctly subsurface deposits. It is initiated after estimating size and value of a deposit by shafts, crosscuts, drifts, raises, and small exploratory holes. Stoping and caving are the two chief plans of operation for underground mining. In stoping, the ore is broken by drilling and blasting with dynamite Caving is generally applied to larger ore bodies with lower metal content. Support is withdrawn from beneath to create a larger caving downfall by a combined use of blasting and breaking the ore or the capping material. Many variations or combinations of stoping and caving are used, but only the most representative are described here.

Metal-mining methods are similar in many respects to coal mining but are more diverse because of greater variations in shape and size of metalliferous ore bodies. The chief difference between the two methods of mining results from the softer character of coal as compared to metallic ores. *See* COAL MINING.

Stoping methods may be classified in several ways but are here considered according to means of ground support. Generally, each method can be further subdivided by the direction of attack in which ore is removed. Thus, overhand stoping involves removal of ore progressively upward from the bottom of the deposit or ore block, whereas underhand stoping removes the ore from the top downward. Overhand methods predominate, principally because both drilling and disposal of broken ore are facilitated by action of gravity

Economic and safety factors are of primary importance in the choice of mining method. These factors involve the strength of wall rocks; strength, shape, dip, thickness, and horizontal area of the ore body; the value of the ore as to distribution, continuity, grade, recoverable value, hardness, friability; cost of the method; degree of adjustability of production and grade to varying metal prices; effects of surface subsidence; requirements for permanent support; availability of filling material; disposition of waste rock; and other, usually minor, factors. Choice of method inevitably involves drilling and blasting procedures, and the method of disposal of broken ore, as integral parts of the mining method. *See* MINING EXCAVATION

Open stopes. Other than occasional stulls, or unframed timbers, for cavities made in the ore body, open stopes use little support. The method is generally limited to steeply dipping, firm ore bodies with uniform character and strong walls. Complete extraction between foot and hanging walls is customary.

Sublevel stoping (Fig. 1) is a relatively low-cost method, distinct from sublevel caving. Haulage levels 100–150 ft apart are connected by raises; sublevel drifts about 50 ft apart vertically are driven from the raise down the center or sides of the ore body to the "end line of the ore block" (underground property division). Bell-shaped feed raises are driven at close intervals from the haulageway to 10–20 ft above. Long-hole or ring drilling is generally employed to blast ore into the bells and collecting fingers which discharge into scraper drifts or directly into cars. Occasionally blasting starts at the end of a lower sublevel and

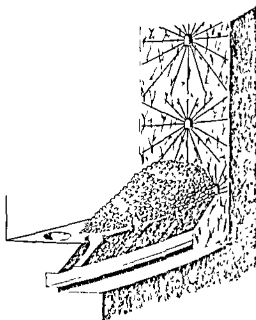


Fig 1. Sublevel stoping

continues in a step pattern upward as the miners retreat in the first sublevel

Glory-holing is an underhand stoping method occasionally used to develop funnel-shaped excavations in large ore bodies. Breaking starts at the top of raises or mill holes driven from level to level and spaced according to the size of the deposit. A slice of ore may be left just under the top level until the stope is mined out A bench is drilled and blasted around the top of the raise, then is widened and deepened in step pattern by underhand methods. The plan view of the raise shows a rough circle or ellipse. The stope must be steep enough for broken ore to flow by gravity to the raise.

Gophering, or coyoting, refers to small-scale mining utilizing small, irregular excavations in ground that generally stands without support. It is employed when capital is not available for systematic development, or where it is the only economical method of mining narrow, erratic veins The term is also used for small, unsupported tunnels in which explosives are placed for breaking large quantities of rock to be used for filling or other commercial purposes.

Supported stopes. Various pillars uphold the openings in supported stopes. Natural pillars of ore; artificial pillars of wood cribbing, masonry, or concrete; broken ore; timber; rock bolts; filling; or combinations of these are used to support the openings in ore bodies and, in some cases, to provide a working floor during mining operations.

Pillar-supported stopes. Columns of ore, or occasionally of wood or concrete, varying in shape and size support the stope. They are splayed out at top and bottom and spaced at intervals depending upon the length of unsupported span desired Such columns are used mostly in flat lying or slightly dipping ore bodies, 10–15 ft thick and with strong overlying rock, but some may be used in much thicker beds The principal advantage in pillar-supported stopes is that mining costs are generally low because of high ore-breaking efficiency and the use of mechanical loading equipment. The disadvantages are loss of ore in pillars and increased danger because of long, and occasionally high, unsupported spans of rock.

Casual pillars of irregular spacing are used where strength of the ground is such that size and spacing can be varied to accord with changing conditions, either of strength or of grade If the ore body is not more than 10–15 ft thick, the method of attack is breast stoping in which the advance is horizontal. In thick ore bodies, a combination method called breast-and-benching is used where the advance is both horizontal and downward in a step pattern, although underhand methods may be used.

Room-and-pillar mining, essentially a more systematic form of the method described above, utilizes square or rectangular pillars left at regular intervals and spaced to form rows. As much as 60% of the ore may be in pillars. These may be left for permanent support or later removed as much as possible; but the latter permits the settling of the roof rock into the mined-out area.

Timber-supported stopes. Such stopes were widely used in the past, but expensive timber is being supplanted in some cases by other methods of ground support.

Square-set stoping. Square sets have been commonly used because of their great flexibility in supporting excavations of irregular outline. The basic unit consists of four upright posts connected in a square at the top by four pieces of timber called caps and girts This unit stands either on solid support such as rock or filling or on other timber. Similar or modified sets may be added laterally and upward (or downward in underhand stoping) to support any shape and size of opening. The ends of the timbers are usually cut in a mortise and tenon pattern so that they fit together snugly for maximum strength Square sets are adaptable either to underhand or overhand methods of stoping, the basic feature being that voids created by ore removal are quickly filled with square sets to support the ground until the remainder of the particular section of the ore body or block of ore can be mined.

The basic weakness of square-set support is that if one member of the timber framework is weakened or destroyed, the entire unit may collapse. This weakness can be overcome in part, either by using rock bolts or by filling the lower section with rock, sand, or tailing which helps to stabilize the square set structure and minimizes the handicap of short timber life under pressure

Stull-supported stopes In narrow veins stulls, unframed timber of varying length and thickness, can be used between the hanging and footwalls for support of open ground.

Filled stopes These are traditionally known as cut-and-fill, although filling may also be used for permanent ground support in square sets, or in shrinkage stopes after the ore has been withdrawn. In cut-and-fill stoping, overhand methods are used, a slice being taken across the back of the ore with the work proceeding from a raise or raises. Temporary support of the open ground, if needed, is provided in the form of timber or, increasingly, rock bolts. When the ore has been removed, generally by mechanical scraping equipment, filling is introduced by a variety of methods to a level close to the back, thus providing permanent support to the walls as well as a working floor for mining the next slice (Fig. 2).

Shrinkage stoping. This is a method utilizing broken ore for ground support and for a working floor during mining operations. Breaking is started a short distance above the haulageway to which the stope is connected by closely spaced chutes. While mining, only enough ore is drawn off through the chutes to accommodate the 40–60% swell of the broken rock. Manways are maintained in the

broken ore at intervals. Packed ore may be loosened by secondary blasting. After extraction of the broken ore, the void may or may not be filled. Size of stopes is limited by the size of the opening which will stand unsupported, although casual pillars or rock bolts may be used for ground support.

Caving. An increasingly used system of mining involves the creation of a void underneath the section of ore body to be mined (undercutting), which allows the ore to cave and break up by its own weight and to be drawn off: or the ore is mined and drawn off from the top, allowing timber matting and the rock or soil capping to cave and follow the mining downward in successive stages. Ore blocks may be square or rectangular in outline and range widely in size. Nominal dimensions are 20,000 to 50,000 ft² in area and 100 to 500 ft in depth.

Top-slicing. In top-slicing, horizontal or inclined slices, 8–12 ft thick, are mined from the top down Timber is used for temporary support. At completion of the slice, the ore below is covered tightly and the posts blasted out, allowing the matting of old timber and the caved capping to rest on the slice below. It is a cheap method with a high ore extraction rate. However, much timber is required, ventilation is difficult, and the fire hazard is high.

Sublevel caving. Adapted from top-slicing, sublevel caving is more efficient, since slices are 15–20 ft thick. The development cost of this method is greater than block caving, but it often provides better selection. Raises at 30–50 ft intervals are driven between haulage levels which are 75–150 ft apart. Sublevels are driven from the raise at vertical intervals of 15–25 ft and horizontal intervals of about 50 ft (in wide ore bodies). Slice drifts are driven normal to the sublevels to the limits of the ore, or to adjacent sublevels, at intervals of about 30 ft horizontally. Mining proceeds from the top down and from the ends of the sublevels toward the raises. The ore over the slice drift is blasted down and usually removed by mechanical scraping. The bottom is lagged or matted, another slice drift driven alongside and the ore blasted down as before. As the capping caves and bears upon the matting, weight is put on the ore pillar below and further mining by combination of the caving action and blasting is facilitated.

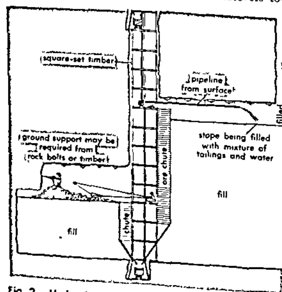


Fig. 2. Hydraulic-fill stope.

Block caving. This is a technique used for mining cheaply large blocks of low grade ore. At the lowest horizon of the block, ore is removed by undercutting and pillar removal. Thus unsupported, the ore immediately above begins to break up from the pressure exerted by overlying ore and capping. As it falls into the undercut area, it is drawn off through a grid of raises on approximately 20-ft centers. When the capping appears at the draw points, drawing is discontinued. When all the ... be resumed from

... ving being

used are gravity and mechanical scraping. The gravity system utilizes pairs of raises driven from the haulage level at intervals of approximately 30 ft and dipping at 60°. At varying vertical distances from the haulage level, usually 30–60 ft, sublevel drifts are driven for screening oversize ore in grizzly screens and for connecting the tops of the raises. Grids are placed over the tops of the raises and, from each one, pairs of small finger raises are driven 15–20 ft. These are belled out (enlarged) on top to receive the ore later to be caved into them. The raises are readily connected at the top by a system of undercut drifts. The pillars between drifts are blasted, thus creating a completely undercut area which initiates the caving process.

In the scraper system, which is rapidly gaining favor, a drift is driven, normal to and just above the haulageway, to the end of the block. A large slusher hoist is installed with the scraper dumping directly into ore cars below. Feeder raises are driven from the slusher drift in pairs at a dip of about 60°. These terminate in sublevel drifts similar to the gravity system or in an undercutting zone. It is general practice to line the slusher and sublevel drifts and raise openings with concrete to better withstand the great weight which they eventually have to bear as caving operations progress (Fig. 3).

Mine shafts. These principal openings into mines—unless adits (or tunnels) are used—should be so located that minimum maintenance will be required. The ideal location is in solid ground in the footwall of ore bodies so that subsequent hanging-wall ground movement resulting from mining operations does not affect the shaft.

Vertical shafts are preferred to inclined because of generally lower sinking, maintenance, and hoisting costs. Many factors are involved, however, and occasionally inclined shafts are sunk to decrease the distance necessary for horizontal crosscuts to reach the ore body.

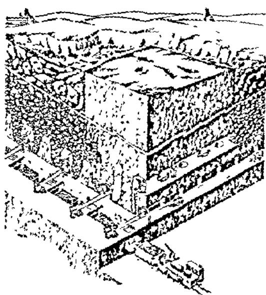


Fig 3. Scraper-drift block caving.

The shape of shafts may be rectangular, square, elliptical, or circular, but the last is coming more frequently into use, due to the maximum strength obtained and the decreased cost and increased speed of sinking. Size depends on the proposed amount of air, ore, supplies, equipment, and personnel to be moved through the shaft. Fluctuating prices of metals, changes in mining methods, and the possible discovery of new ore bodies require that the initial design be made with careful consideration of future requirements.

Shaft sinking in rock. Increased mechanization has increased speed and reduced development time necessary before the production stage in mining. In 1957, a record advance of 834 ft in 30 days was made in a 30-ft diameter circular shaft in South Africa. It is probable that 1000 ft per month will be reached. Chief elements in sinking a shaft are drilling and blasting, mucking, hoisting the broken rock, and establishing ground support. Where water is a serious problem, expensive installation and maintenance of large pumps have been eliminated by drilling holes down and to the side, then filling them under high pressure with a mixture of cement and water which penetrates fractures and joint planes to decrease the flow of water toward the shaft. This is called cementation.

Shaft sinking in soft ground. In soft ground shaft sinking is generally confined to parts near the surface. Various methods used include: (1) wood or steel spiling (forepoling) driven on the perimeter and ahead of the shaft bottom, whereupon the ground is removed and permanent wall support installed; (2) sinking the shaft lining of wood, steel, or concrete into the soft ground before shovel or clamshell removal of earth; (3) forcing grout into the unconsolidated ground; or (4) freezing the ground by circulating refrigerant through pipes penetrating the ground.

Sinking equipment. Included in sinking equipment are hoist, headframe, pumps, fans, buckets for the broken rock, dumping facilities on the surface, and often an operating stage or platform from which is suspended drilling and mucking equipment. Lining and equipping of the shaft may also be done from the platform. Services needed include electricity, compressed air, and water. Permanent services may or may not be installed during sinking operations. *See* MINING OPERATING FACILITIES

Support of mine workings. Most rock formations require support of underground workings, although narrow drifts, small shafts, and even larger openings will stand for many years in some types of rock. The amount and type of support required depends on depth, the time ground must remain open, and the character of the rock.

Subsidence and deep-level rock flow are movements of ground toward and into voids created in underground mining. The angle from the surface taken by subsidence depends on the character of the rock or upon planes of weakness such as faults; but usually it assumes about 63°.

Types of support vary in the degree to which they yield under pressure. It is generally desirable to prevent the start of ground movement because, once begun, it may seriously impede subsequent mining, above or adjacent.

Pillars. These are sections of the ore body left temporarily or permanently for ground support or for water or ventilation seals. The principal uses as support are around shafts, under surface installations subject to subsidence, in some types of stoping, and in connection with any mine opening that it is desired to retain. Relatively incompressible, pillars are an excellent means of support. However, they may lead to sudden caving from great pressures set up around them or because of rockburst hazards in brittle rocks. Their use may also result in irregular subsidence on the surface with considerable damage to surface installations. Pillars may or may not be mined out later depending on their value and the hazards involved in their extraction.

Timber. Historically timber has been the principal means of supporting mine openings because of its former low cost, general availability, and flexibility in use. Since it is compressible, thus allowing ground movement to start, and subject to decay, it is used most effectively only for short-term support or in areas where ground movement is of small magnitude.

To inhibit decay, timber may be impregnated under pressure with creosote, zinc chloride, or other preservative compounds.

Rock bolts. Rock bolts increasingly supplant timber in many mine support usages. They provide tensile strength to rock which is strong only under compression. The most common types are fabricated of round steel rods ¾–1 in. in diameter and 4–8 ft long. One type has threads on one end and a slot in the other. It is anchored by driving the

split end of the bolt onto a steel wedge at the bottom of a drill hole of suitable diameter. The resulting expansion locks the bolt in the hole. A bearing plate of wood or steel on the threaded end is tightened to the rock surface by means of a nut. The other type of bolt has a square head on the outer end and is anchored by screwing the threaded, inner end into an expansion shell assembly that locks the bolt against the walls of the drill hole. Rock bolts are safer, more economical, and easier to use than timber, allow larger open areas for unobstructed use of mechanical equipment, are adaptable to any shape of excavation, and retard the start of ground movement.

Filling. Mined-out areas are commonly filled to prevent or decrease ground movement, to extinguish or prevent fires, and to improve ventilation. Waste rock, sand, gravel, slag, tailing, or combinations thereof are used for this purpose. Filling placed dry contains a high percentage of voids which permits later compaction and ground movement. When filling is placed hydraulically and then drained, void space and subsequent compaction are greatly reduced. Tailing emplaced hydraulically is generally the most satisfactory filling.

Steel sets. Steel is increasingly used where large-scale rock movement is encountered and timber or rock bolting is inadequate. H beams in 4- to 12-in. sizes are most commonly used. The form may be that of two legs and a cap, straight legs and an arched top, or a circle. The sets are covered with lagging which fails under pressure and can then be replaced, thus protecting the steel sets from a constantly increasing pressure which would cause buckling. Although initial costs are greater than timber, maintenance under heavy ground pressure is less.

Yieldable arches. These are made of steel in modified U shape. The segments, commonly three, are held together with U bolts which allow them to slip when pressure overcomes the frictional resistance. Their success depends on yielding under load until the rock forms a natural arch and movement ceases. The three-piece set is not always satisfactory in loose, unconsolidated ground, or where differential pressures are exerted, as in a fault zone. Circular sets may overcome these disadvantages.

Concrete. In many ways concrete is the most satisfactory support for mine openings. It is used increasingly in monolithic, precast, or prestressed forms for shaft linings, and support members, ore pockets, stations, tubes, dams, bulkheads, and grizzly and slusher levels in cave mining. It largely prevents the beginning of ground movement and does not fail until ground pressures become exceedingly great. The capital cost for equipment to mix and transport concrete is high, but maintenance costs are usually low compared to other forms of support in very heavy ground. [E.R.B.]

Bibliography: C. F. Jackson and J. H. Hedges, *Metal Mining Practice*, U.S. Bureau of Mines Bulletin 419, 1939; R. Peele, *Mining Engineer's Handbook*, vol. 1, 3d ed., 1950; B. Stoces, *Introduction to Mining*, 2 vols., 1954.

Mining excavation

In mining for coal, metallic, and nonmetallic minerals, excavation is the process of breaking minerals from their solid position and loading and transporting them to the surface. It includes (1) use of explosives and detonators, (2) distribution of explosive through rock by means of a pattern of shot holes, (3) explosive and mechanical fragmentation, (4) loading, and (5) transportation. Excavation is also involved in establishing mine entries and other development workings in waste rock for access to the minerals.

All hard rock is broken with explosives to attain fragmentation. Some moderately soft deposits such as coal, potash and borax, are fragmented directly by machines without explosives. When fragmentation is by machine, the loading device is commonly an integral part of the machine.

Transportation is involved in mining all mineral formations whether at the surface or underground. Although gravity, requiring little or no machinery, is employed in transporting broken material from a higher to a lower elevation, generally a wide variety of machines and equipment is required.

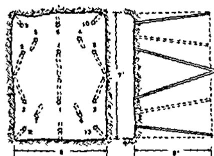
Explosives. Two general classes of explosives are used for mining: blasting powder and high explosives, or dynamites. Black powder is no longer used in underground mining as it is both unsuited for breaking hard rock and unsafe. The base for most high explosives used in underground mining is nitroglycerin, a compound of carbon, hydrogen, and nitrogen, which is explosive in itself. When detonated, it is decomposed into carbon dioxide, water, nitrogen, and oxygen at an increase in volume at atmospheric pressure of 1000-fold. Ammonium nitrate is replacing nitroglycerin to a considerable extent in dynamite for underground blasting. Mixed with a hydrocarbon, it is used extensively for open-pit blasting.

Many other ingredients are used in making high explosives, but most of them are either oxygen carriers (sodium nitrate) or combustibles (wood meal). "Permissible" explosives (allowable under U.S. Bureau of Mines specifications), when used in blasting coal, contain in addition ammonium or sodium chloride to reduce the temperature of detonation for safety.

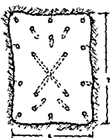
High explosives are detonated with blasting caps—small metal tubes or shells closed at one end and charged with a highly heat-sensitive explosive.

Fragmentation. Fragmentation is the process of breaking ground with explosives and machines.

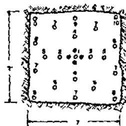
Explosives fragmentation. The object of explosives fragmentation is to break the minerals and produce fragments of a size best suited for handling. It is cheaper and generally desirable to break waste rock in coarse sizes. Customarily,



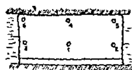
HORIZONTAL V CUT

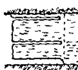


4-HOLE PYRAMID CUT



5-HOLE BURN CUT





UNDERCUT COAL FACE

Fig. 1. Typical rounds for explosives fragmentation.

coal has been produced as lumps, but now the trend is toward using fine sizes for more economical handling and combustion. Fine sizes are desirable for most metallic and some nonmetallic ores that are processed after mining. Some nonmetallic minerals can be marketed more profitably in medium coarse sizes (¼–2 in.), so fine material is wasted and must be minimized. The size of mineral fragments can be controlled partly by the amount and strength of the explosive used in blasting and somewhat by the spacing between shot holes and by their depth.

An assemblage of shot holes drilled into the face of a stope, drift, crosscut, shaft, raise, winze, adit, or tunnel and blasted at one time is a "round" (Fig. 1). The pattern of the round contributes to the effectiveness of a blast. A wide variety of patterns is used, depending upon the character of the material to be broken, the size and shape of the desired opening, and the desired size of fragments. These rounds have been given names such as pyra-
mi
of
cut
wegian and Michigan, have been named according to place of origin. Michigan is also called burned cut because the holes which are drilled close together in a group near the center of the face, when blasted, pulverize the rock and discharge it at a high velocity. *See* BORING AND DRILLING, MINERAL.

To break a round, each hole is charged with several cylindrical cartridges of dynamite. A detonating cap with attached safety fuse or electric wire is inserted into one of the cartridges in each hole to form a primer. The round is then blasted by igniting the fuse or sending an electric current through the wiring. The sequence in which holes of a round are blasted is important. Usually a group of holes near the center or along one side is blasted first. Then successive groups of holes are blasted in series. Crude timing between successive blasts is attained by trimming fuses to different lengths. Precision timing is possible with electric blasting caps constructed to vary in detonation by a few thousandths of a second.

Mechanical fragmentation. This method is used primarily in coal mining. The principal machines are the continuous miner, auger, coal plow, and coal cutter. The continuous miner is manufactured in several sizes by both domestic and foreign companies. There are many variations in detail to meet varying conditions in coal seams.

One representative model is shown in Fig. 2. The coal is fragmented by a front-end cutting head comprising a number of continuously revolving chains upon which are mounted hard-metal-tipped cutters or picks. The chains are mounted on a bar which can be rotated horizontally to cut an entry 12–20 ft wide and vertically to cut from 6 in. below the bottom of the machine to 7 or 8 ft above. The cutter-bar frame is mounted on a carriage which in turn is mounted on caterpillar treads. The broken coal is collected by a conveyor which transports it from the cutting head to the transportation system at the rear of the machine.

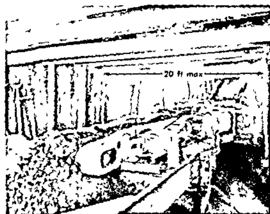


Fig. 2. A representative 1958 model of a continuous miner finishes cut as it rips overlapping paths in coal seam 18 in. deep by 42 in. wide by 90 in. high. Coal is discharged into waiting shuttle. (U.S Bureau of Mines and Joy Mfg.)

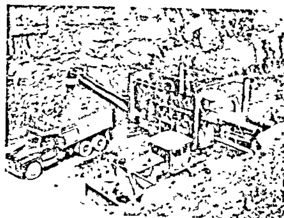


Fig. 3. View from spoil bank showing complete operation of modern auger mining and coal being loaded into truck in the Pittsburgh coal bed of West Virginia. (*U.S. Bureau of Mines*)

Augers up to 5 ft in diameter (Fig. 3) are used under high banks of abandoned strip mines to recover coal to a boring depth of about 200 ft. The auger is made in sections, the first one having a hard-metal-faced cutting tool that does the cutting as the auger is rotated. The force providing thrust and rotation to the auger is supplied by a diesel engine at the surface Some machines are self-propelling and handle and store the auger sections mechanically.

The coal plow or planer was developed in Germany to mine coal by the longwall system. The plow is drawn back and forth along the face and cuts or slabs off the coal in a series of slices. The planer unit comprises (1) an armored double-chain (Panzer) conveyor which rests on the mine floor and extends the entire length of the longwall face. (2) an electric driving unit at each end of the conveyor, (3) the planer assembly, and (4) pneumatic conveyor shifters. The electric motors move the plow by means of a rotating drum and chain attached to the planer. Broken coal falls to the conveyor which moves it to entries adjacent to the mining areas. The pneumatic shifters move the conveyor and the plow toward the face as mining advances.

Where coal is fragmented by explosives, it is first undercut by a coal-cutting machine along the bottom of the seam. It either falls by gravity (undercutting and mass falling), or is broken by permissible explosives or other substitutes. Shot holes are placed to take advantage of such natural features as the presence of hard and soft bands and the direction of cleavage planes in the coal.

Three classes of machines are used for cutting coal: disk, bar, and chain, classified by the method employed for holding the cutting tools or cutter picks.

The chain coal cutter (Fig. 4) is in most common use and is manufactured in various sizes. These electrically powered machines are designed to cut in an arc of 180°. The smaller units are maneuvered by power-actuated drums and cable attached to stationary anchors. The larger self-propelled units are mounted on track or use pneumatic tires. The cutting jib can be rotated without moving the machine while cutting. It can also be rotated to cut either vertically or horizontally.

Loading. Since most loading is mechanized, hand shoveling is limited. Underground loaders include the continuous loader, scraper loader, and revolving and overcast shovels. Clamshell loaders are used in sinking shafts. The continuous loader (Fig. 5) is used extensively in underground coal mining and less extensively in other mines Its essential parts comprise a gathering head frame, or scoop, on the front end that is crowded into the broken material by the machine and a set of gathering arms to rake the material onto a bar chain conveyor that transfers it to the haulage system. The scraper loader consists of a hoe-type scraper, double-drum electric hoist, and ramp. Fragmented material is scraped up the ramp and discharged into the transportation system. The scraper loader is used in many places in mining to scrape ore and waste short distances up to about 200 ft. Revolving power shovels are used for excavation in large un



Fig. 4. Cutter making a shear cut in coal face of a West Virginia coal mine. (*U.S. Bureau of Mines and Joy Mfg.*)

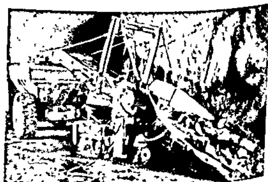


Fig. 5. Caterpillar-mounted electric machine loading broken limestone into diesel-powered haulage truck in deep limestone mine in Ohio. (*U.S. Bureau of Mines*)

derground openings. The overcast shovel loader, which runs on rail or caterpillar tracks, is best adapted for use in confined areas. Loading is accomplished by crowding the unit into the broken rock; when the shovel is full it casts back over the unit into cars or trucks.

Transportation. Moving or hauling of men, supplies, and broken minerals is one of the most complicated operations in mining. Transportation includes: (1) transfer of broken material through mine openings by gravity and scraping, (2) rail haulage on surface and underground, (3) trackless, wheeled haulage on surface and underground, (4) hoisting and cable haulage from open pits and underground mines, (5) movement of broken material by numerous types of conveyors, and (6) pumping broken ore through pipelines.

Gravity and scraping methods are used to gather small quantities of material into larger volume for transportation by some other method. Broken material may fall directly into raises and ore passes where it flows by gravity to a chute or ore pocket.

Underground rail haulage has been adapted from railroad practice, but on a much smaller scale. It is used mostly for collecting broken material from chutes (Fig. 6) and various loaders, and transporting it to the hoist or directly to the surface. Track gage ranges from 18 to 48 in. Trains are made up of cars ranging from 1 ton to about 20 tons capacity. Cars are equipped for manual dumping or with some dumping device. Small units are drawn by storage battery locomotives and larger units by [illegible]

cars with self-contained discharging conveyors (Fig. 7). They are mounted on pneumatic tires and have capacities ranging to 20 tons. Dump trucks are used underground to a limited extent. When so used they are equipped with diesel engines for



Fig. 6. Loading 60-ft³-capacity, Granby-type mine car at uranium ore-pass station on haulage level, San Juan County, Utah. Underswing arc gate is operated by a compressed-air cylinder controlled by motorman helper on right. (U.S. Bureau of Mines)



Fig. 7. Loader and shuttle car in a Missouri lead-zinc mine. (U.S. Bureau of Mines)

safety. Most trackless underground haulage equipment negotiates grades up to 10%.

Electric hoisting is employed in deep underground mines. Either vertical or inclined shafts are divided into three to six parallel shaftways—one for a manway, pipe, or power lines, and a minimum of two for hoisting. In large deep shafts, two compartments are commonly used for hoisting broken material, and two others for handling supplies, equipment, and personnel.

The hoisting layout comprises a headframe, usually of timber, concrete, or steel, erected over the collar of the shaft. A sheave wheel is mounted at the top of the headframe for each hoisting compartment. In addition, the headframe contains bins into which hoisted material is discharged.

Most hoists are electrically driven and have two [illegible]

range up to about 1½ in. in diameter. Rock is hoisted from various levels of the mine in skips (buckets, baskets, or open cars) ranging up to 10 tons capacity. Men and supplies and, at some smaller mines, loaded cars are hoisted in cages with one to three decks.

An important safety feature of the skip and the cage is a device that stops them if the hoisting cable fails. When failure occurs, a spring releases a set of safety catches that engage the wooden guides along each side of the skip or cage, thereby stopping it.

Several different types of conveyors are used in transporting coal in underground mines. The most important of these are belt conveyors. In some mines virtually all the coal is transported by a system of belt conveyors from the mining face through the various entries to surface. [P.T.A.]

Mining machinery

Apparatus used in excavating rock and loose surface material, usually from a greater depth, in greater volume for a given surface area, and fo

longer period of time than is common in construction. In general, mining machinery also includes a wide variety of equipment used for finding (prospecting and exploring), removing (mining, developing and exploiting, or excavating), and improving (processing, milling or beneficiating, and refining) valuable constituents of the earth. Mining machinery may also include metallurgical (smelting) and chemical processing equipment used for further separating and extracting products of mineral origin.

Design and construction. All components of mining machinery—including mechanism, controls, means of powering, and frame—require the following features to a much greater extent than do other machines (possibly excepting some military, construction, oil well, and marine units).

Ruggedness. Equipment is handled roughly, frequently receiving severe and sudden shock from dropping, striking, and blasting vibration; overloading is common, and long life is demanded by the economics of mining.

Weather resistance. Operations extend over a wide range of climate and altitude.

Abrasion resistance. Minerals include some of the hardest substances known. Dust and fine particles are always present.

Water and corrosion resistance. Moisture and water, often acidic, are common in mining operations.

Infrequent and simple maintenance. Equipment is often widely scattered and in locations with restricted access. Trained mechanics and repair parts are generally limited in availability because of the remoteness of operations.

Easy disassembly and reassembly. Access to the machinery at the site of operation is frequently limited. Also the working space near it and mechanical aids to moving or lifting it may be limited or nonexistent.

Safety. Mining has had a poor safety record. As a result, most governments have testing bureaus, inspection agencies, and enforcement laws for the approval of mining equipment.

Simplicity of operation, low initial cost, and low operating costs are also desirable features of mining machinery.

Underground requirements. Machinery operated underground must meet special design requirements.

Low ventilation demand. Quantity and geometry of air courses are fixed rather closely so that high air consumption is a problem and noxious gases cannot be readily dispersed. Heat removal is a problem in deep mines.

Compactness. Space is at a premium, especially hei[illegible]

[illegible]

onl[illegible]

lan[illegible]s.

Hand portability. Units or components must frequently be hand carried into an operating area.

Absence of spark and flame. Equipment is often used in or near explosives, timber supports, and natural or man-made combustible gas or dust. In the presence of hydrocarbons as well as of certain metal ores such as some sulfides, complete absence of open sparks or flames is a major requirement.

Exploration machinery (vehicles, drills, and accessory equipment) is subjected to the same operating conditions as other mining equipment. Machinery for beneficiating must have, above all, abrasion resistance. Smelting equipment has the requirement of heat resistance. Chemical refining process equipment must be highly resistant to corrosion. Reliability of processing equipment is critical because slurries are commonly processed, and they can cause considerable difficulty in restarting after shutdowns.

Power source. Mining machinery is most commonly operated by compressed air. Electricity is also widely used and is often the basic source. Compressed air has the advantages of simplicity of transmission and safety under wet conditions. It aids ventilation which makes it especially advantageous underground. Above ground, compressed air is useful largely because machines powered by it yield more readily upon overload or jamming. Large central compressors are common, especially at underground mines.

Electric power, purchased or generated at a large central station, is probably more common for open-pit and strip mines and for dredging operations. Underground coal and saline mineral mines use considerable electric-powered production machinery, but at other underground mines electricity is normal only for powering pumps and haulage units in relatively permanent locations. Direct current devices dominate the field, but alternating current apparatus is rapidly being adopted. Many units are battery or cable-reel types. Processing machinery is almost exclusively powered by electricity.

Diesel engines are becoming popular for local power and especially on haulage units. For underground application of diesels, abundant ventilation is essential even if wet scrubbers, chemical oxidizers, and other accessories are used to aid the removal of noxious and irritant exhaust gases. Hydraulic (oil) control and driving mechanisms are widely used. Devices actuated by wire rope are common, too.

Mining machinery includes a wide range of excavating and building equipment such as that used in the construction industry, and processing equipment such as that used in the metallurgical and chemical industries. Unique mining machines include drills, articulated cutters, specially adapted materials handlers, and mine site processers.

Drills. Three principal functions of drills are to obtain cores during exploration, to prepare blast holes, and to extract ores. Drills for exploration are of various diameters; drills for blasting generally produce small- to medium-diameter holes;

drills for extraction wells tend toward large diameters.

Exploration drills Rotary coring action of a diamond-studded bit is used in exploration to penetrate the hardest rocks. Large units make holes under 3 in. in diameter and 5000 ft or more deep; at the other extreme are units so small that they can be pack-carried.

Water-jet drills. For exploration and for blasting holes in loose or weakly bonded materials, water-jet drills are used.

Blasthole rock drills. Percussion, rotary, and combined rotary-percussion actions are used for holes up to 12 in. or more in diameter by 125 ft or more long from the surface, and 1–3 in. by 5–200 ft from underground. Crawler or wheeled carriers are used, and the smaller drills are often mounted on hydraulic booms.

Jet flame blasthole drills. For surface work in hard abrasive rock a flame drill is used to make blast holes economic.

Articulated cutters. Mining is becoming more mechanized. Mechanical mining machinery takes various forms, all directed toward continuous removal of ore.

Augers. Coal and soft sediments are mechanically mined by augers up to 5 ft in diameter and 100 ft long.

Corers and full-face borers. For vertical and inclined openings up to 8 ft in diameter and several hundred feet long, corers and borers are used. These machines can dig tunnels of 26-ft diameter and of infinite length, and can be used in medium-hard sediments.

Continuous-miners. For tunnel-like openings in coal and saline deposits, continuous miners use lugged chain and lugged drums to rip, and conveyors to move the material.

Planers. In coal and mineral deposits of medium hardness planers break off 6-in. thicknesses from long walls 3–5 ft high.

Rock saws. Channelers use diamonds and other abrasives and also flame jets to remove material.

Hydraulic monitors. Water nozzles up to 6 in in diameter operated at pressures of several hundred pounds per square inch are used to excavate partly cemented surface material and brittle hydrocarbons both on the surface and underground.

Materials handlers. After the ore is dislodged from its site, it is moved to a common loading point for transportation to mill or refinery. The moving equipment used at the mine is classed as materials-handling machines to distinguish it from transportation equipment, which moves the material over greater distances. Excavating machines are essentially materials-handling units, and materials-handling machines may be equipped with self-loaders that excavate.

Excavator loaders. Broken ground is removed and loaded in limited space by various grab-bucket shaft muckers, overcasting shovel tunnel muckers, low gathering-head loading conveyors with eccentric arms, lugged chains, screws, or shaking bills. Dragline scraper slushers perform similar functions.

Dipper shovels and dragline cranes. From ¾ to 70 yd³ of material can be moved in a load by shovels or draglines on crawlers or walking shoes.

Bucketline and bucketwheel excavators. Over 5000 yd³/hour are moved by multibucket devices on crawler, railroad, or floating hull mounts.

Suction pump floating dredges. Capacities of dredges range up to 3700 yd³/hour.

Trucks. Electric trucks and trailer trucks of low height, with conveyor bottoms, have capacities up to 80 tons.

Railroad locomotives. Underground locomotives range from ½ to 80 tons in weight. They are powered by electric storage battery, trolley wire, trailing cable, or compressed air.

Conveyors Movable, self-propelled, sectional, and extensible conveyors use belts, aprons, chains, or buckets.

Wire rope hoists. Winders rated up to 6000 hp for shafts 6000 ft or more deep lift material to the surface.

Auxiliary machines. Safe, efficient mine operation requires such auxiliary equipment as fans and grinders Pumps with heads of 500–1000 ft drain mines; other pumps, for dredges, handle 6-in. solids. Mechanically moved roof support bars have capacities up to 100 tons. Fans may have ratings of over 100,000 ft³/min. Crushers and grinders with openings up to 5 ft handle soft salt or such extremely hard material as iron ore To separate minerals by size or weight, screens, classifiers, and concentrators operate by vibration, fluid flow, or induced centrifugal motion. Other concentrators use properties such as physical or chemical froth flotation, magnetic attraction, or electrostatic separation. See BULK-HANDLING MACHINES; GRINDING MILL; MINING OPERATING FACILITIES. [L.E.AN.]

Bibliography: Annual Mining Guidebooks; R. Peele (ed.), *Engineering and Mining Journal; Mining Engineers' Handbook,* 2 vols., 3d ed., 1941; U.S Bureau of Mines Publications.

Mining operating facilities

Physical aids, procedures, and plans widely used to expedite development and production in mining. Several of these are of such outstanding significance that they are the subject of separate articles (see BORING AND DRILLING, MINERAL; MINING EXCAVATION). This article includes discussions of seven other facilitating aspects of mining: power, ventilation, illumination, drainage, storage and shipping, field sampling and ore estimation, and mine evaluation.

MINING POWER

Power is applied to mining in six distinctive ways: electricity, diesel power, compressed air, hydraulic power, steam power, and hydroelectricity.

Electricity. Both alternating and direct electric current are used in modern mining. Direct current (dc) is used for locomotive haulage and for the major portion of underground coal operations because of the high torque developed in dc series motors under heavy loads and starting. Alternating current (ac) is used less extensively although some mines employ both in a dual system.

In the United States, 6000-, 4160-, and 2300-volt three-phase alternating current is generally received from service companies at the surface substation. In a well-planned mine, all the power is metered at one point and the power factor is adjusted between 90 and 95% by the use of synchronous motors and capacitors. Neoprene-covered cables, type SHD, 15,000 volts are recommended to take the power underground through boreholes or power shafts. Each of three insulated conductors is covered with copper shielding braiding to eliminate static stresses, and a ground conductor for each power conductor is placed in the interstices of the cable.

Converting ac to dc underground may be accomplished by several methods. A 250- or 500-volt substation can be provided by converters, motor-generator sets, mercury-arc rectifiers including the glass-bulb type, and dry rectifiers made of selenium, germanium, or silicon. Portable conversion units offer increased convenience and efficiency.

Trolley and feeder lines supply power throughout the main roadways. These distribution lines should not be extended more than 3500–4000 ft from the power source to avoid low voltage at the end of the line. If local laws permit, 500-volt systems are sometimes used to minimize line voltage drop. This supplies twice the load for the same size and length of cable, or conversely, the same load can be supplied at twice the distance. Trailing cables, fastened to the trolley nipping stations or power centers, supply power to machines in sections of the mine where explosive gases may be present.

Sectionalization is a method of distributing mine power so that power cables can be isolated for reasons of faults or repairs without shutting off the main supply to several working sections. For main distribution, circuit breakers, disconnect switches, and various overcurrent protective devices are essential for properly installed sectionalization. At face areas, safety circuit centers and associated intrinsically safe circuitry make it possible to connect or disconnect short cables without danger of causing incendiary arcing which could ignite gas.

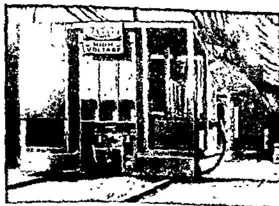


Fig. 1. A 300-kw Westinghouse ignitron rectifier installed at Willow Grove Mine, Hanna Coal Company.

Alternating-current power for mining is increasing in popularity because its equipment is less costly than dc and its maintenance is simpler. Alternating-current motors, for example, cost one-third to one-half as much as the equivalent dc and are more compact. High voltage is transformed to 440 or 220 volts at underground substations. Portable units are also utilized. Sectionalization is applicable to ac distribution.

Alternating-current power is commonly used in strip mining. High voltage of 33,000 volts is stepped down to 7300, 6600, 4160, or 2300 volts. Permanent substations equipped with lightning arrestors, circuit breakers, ground protective equipment, and other protective devices, and semiportable substations are employed for distribution. Power is distributed by pole lines or cable systems, or a combination of the two.

Diesel power. This type is rapidly gaining favor in metal and noncoal mining because of its flexibility. Some states require that underground diesel-powered equipment be approved by the U.S. Bureau of Mines. Details of these standards include explosion-proof housings, mine ventilation necessary to dilute exhaust gases, control of hot exhaust gases and surface temperatures, concentrations of toxic constituents in exhaust gases, and recommendations for the selection, handling, and storage of diesel fuel oil.

Compressed air. As a mine facility, pneumatic power is utilized in a variety of applications, mostly in the metal and noncoal fields. It is used to power drills and hoists; pneumatic tools, such as grinders, drills, riveters, chippers, pneumatic diggers, and spades; air-driven sump pumps; direct acting and air-lift pumps; pile drivers for shaft sheathing; air pistons for unloading cars; drill-steel sharpeners; air motors; compressed-air locomotives; shank and detachable-bit grinders; mine ventilation; and in supplying air for blowing converters; starting diesel engines; and coal preparation. Compressed air is used in coal mining for blasting. This method works with 9000–10,000 pounds per square inch (psi) and the air is released from a tube which is inserted in a hole in the face of the coal when a metal diaphragm ruptures. The force is released through slots in the tube and breaks up the coal which had been previously undercut.

Hydraulic power. For mining, hydraulic applications are rather limited in usage and may employ either water or oil as a fluid. Oil types are used in connection with small tools, lifts, and in the intricate operation of continuous mining machines and

other equipment. An available waterfall may be directed to power equipment such as air compressors. A unique use of hydraulic power, called jet mining, uses air pressure to force water through ¼-in. nozzles under 2000 psi pressure. This jet action has been developed to cut a material called gilsonite, a solid hydrocarbon, by use of water at a rate of 300 gallons per minute.

Steam power. Although formerly used to drive compressors, hoists, generators, and other equipment, steam is now rarely used in mining but has definite, although limited, application in some coal mines in which there is an abundance of waste fuel or unmarketable coal. Parts of certain coal seams contain impure bands that must be rejected, or that are difficult to clean, but they can be burned under boilers with special firing equipment.

Hydroelectricity. As applied to mining, hydroelectricity is used mostly in the electrometallurgical fields, where vast amounts of power are required at a reasonable cost. Some hydroelectricity has been used for normal, electrical, mining power, but mostly it is used for processes beyond the mining operation such as beneficiation, smelting, and refining. In certain cases these are done near the mine mouth in isolated areas in order to reduce bulk before shipment. [R.S.J.]

Bibliography: *Coal Age*, Mining Guidebook and Buying Directory Issue, 61(7a):70, 1956; M. A. Elliott and R. S. James, *The Development of Permissible Requirements for Safe Underground Diesel Haulage*, U.S. Bur. Mines, Inform. Circ. 7673, 1953; R. Peele (ed.), *Mining Engineer's Handbook*, 3d ed., 1941; Vertical-seam mining by jet cutting, fluming, and pumping, *Coal Age*, 62(11):80–83, 1957.

MINE VENTILATION

The purpose of mine ventilation is to provide comfortable, safe, and healthful atmospheric conditions at places where men work or travel.

Air-flow fundamentals. The following three points summarize air-flow principles for mine ventilation: (1) air flow is induced by a pressure difference between intake and exhaust; (2) the pressure created must be sufficient to overcome the system resistance and may be either negative or positive; and (3) air flows from the point of higher to lower pressure. Also, mine air flow is considered turbulent and follows the square law relationship between volume and pressure, that is, a doubled volume requires four times the pressure. See VENTILATION.

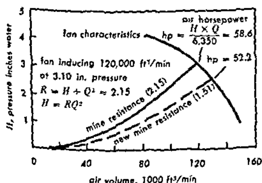


Fig. 2. Graph of fan characteristics versus mine resistance.

The principles of fans may be summarized as follows: (1) air quantity varies directly as fan speed, and is independent of air density; (2) pressures induced vary directly as the square of the fan speed, and directly as the air density; (3) the fan power input varies directly as the cube of the fan speed, and directly as the air density; (4) the fan mechanical efficiency is independent of fan speed and density. See FAN.

The amount of air movement induced will be dependent upon the fan characteristic and mine resistance as shown by Fig. 2. The pressure H required to pass a quantity of air Q through a mine or segment is expressed by the formula $H = RQ^2$,

expressed without decimals, and the quantity Q of air in cubic feet per minute (cfm) should be di-

What pressure is necessary to induce 60,000 cfm through a single airway 6 ft high, 12 ft wide, and 2500 ft long in sedimentary rock? The airway is straight, has average irregularities, and is moderately obstructed. K from table is 70; l is 2500; perimeter is 36; area is 72 ft². Q is 60,000 cfm ÷ 100,000 = 0.60.

$$\text{Pressure} = H = RQ^2 \qquad R = \frac{KlO}{5.2A}$$

$$H = \frac{KlOQ^2}{5.2A^3} = \frac{70 \times 2500 \times 36 \times (0.6)^2}{5.2 \times (72)^3}$$

$$= 1.17 \text{ in. water pressure}$$

Parallel air flow can be determined by the square-law relationship. For example, the pressure H required to pass 60,000 cfm through 2500 ft of single entry is 1.17 in. of water. For two identical entries the pressure is

$$H_1\left(\frac{1}{2}\right)^2 = H$$

$$H = \frac{H_1}{4} = \frac{1.17}{4} = 0.292 \text{ in. water}$$

When resistance factors are known or entries are not identical, the formula is

$$\frac{1}{\sqrt{R}} = \frac{1}{\sqrt{R_1}} + \frac{1}{\sqrt{R_2}} + \frac{1}{\sqrt{R_3}} + \cdots + \frac{1}{}$$

For example, what is the combined resistance factor of two entries 1000 ft long; one entry is substantially larger, $R_1 = 1.50$, $R_2 = 4.0$. Also, what pressure is required to pass 60,000 cfm through 2500 ft of the combined entries, assuming average conditions throughout?

$$\frac{1}{\sqrt{R}} = \frac{1}{\sqrt{1.5}} + \frac{1}{\sqrt{4.0}} = \frac{1}{1.22} + \frac{1}{2.00}$$

$$= 0.82 + 0.50 = 1.32$$

$$\frac{1}{\sqrt{R}} = 1.32 \qquad R = 0.58 \text{ per 1000 feet entry}$$

$$R \text{ for } 2500 \text{ ft} = 2.5 \times 0.58 = 1.45$$

$$\text{Pressure} = 1.45 \times (0.60)^2 = 0.52 \text{ in. water}$$

Air quantity requirements. Unless covered by state laws, the common criteria for adequate ventilation are absence of smoke and dust with moderate air temperatures in metal mines and the absence of methane, smoke, and dust in coal mines. Natural conditions of gas, rock temperatures, dust, and operating practices determine requirements. Good quality air is not deficient of oxygen and is free of harmful amounts of physiological or explosive contaminants *See* MINING SAFETY.

Mine gases. Important contaminants of mine air are carbon dioxide, hydrogen sulfide, methane, carbon monoxide, and sulfur dioxide.

Carbon dioxide, CO_2, specific gravity 1.529, is produced by oxidation and combustion of organic compounds and is occluded in the rock strata of certain mines. It is heavy, colorless, and odorless and is [illegible]

H [illegible]

the [illegible]

It is colorless, has the odor of rotten eggs, and may be found near areas of stagnant water in poorly ventilated areas.

Methane, CH_4, specific gravity 0.554, is a natural constituent of all coals. It may be occluded in carbonaceous shales and sandstones, and may infiltrate into metal mines at contacts with carbonaceous rocks. It is colorless, odorless, and may be found in high, poorly ventilated cavities.

Carbon monoxide, CO, specific gravity 0.967, is not a normal constituent of mine air, but is produced in mines by the incomplete combustion of carbonaceous matter, mine fires, or from gas or dust explosions. It is colorless and odorless.

Sulfur dioxide, SO_2, specific gravity 2.264, is not common, but may be found in sulfur mines and in mines with rich sulfide ores as the result of fires. It is a water-soluble and colorless gas with a suffocating odor.

Black damp is a common term applied to oxygen-deficient atmospheres; it is not a specific gas mixture but may contain any of many gases produced by oxidation and processes that use oxygen and liberate carbon dioxide.

Small quantities of air contaminants must be determined by laboratory analysis of air samples. On-the-spot safety determinations for carbon dioxide and oxygen deficiency may be made by flame safety lamp or small portable absorption instruments. Methane can be detected by flame safety lamp and commercial testers; carbon monoxide and hydrogen sulfide by special hand-held colorimetric indicators. *See* SAFETY LAMP.

Dust and dust hazards. Dust is defined as the solid particulate matter thrown into suspension by mining operations. The size of particles may range upward from less than 1 micron (0.001 mm) diameter (as shown on Table 2, the Frank chart); particles larger than 10 μ can usually be seen by the naked eye. The dust hazard may be physiological or explosive, or both physiological and explosive as is

Table 1. Reasonable frictional coefficients

		Straight			Sinuous or curved: Moderate			Sinuous or curved: High degree		
Type of airway	Irregularities	Clean	Slightly obstructed	Moderately obstructed	Clean	Slightly obstructed	Moderately obstructed	Clean	Slightly obstructed	Moderately obstructed
Smooth-lined	Minimum	10	15	25	25	30	40	35	40	50
	Average	15	20	30	30	35	45	40	45	55
	Maximum	20	25	35	35	40	50	45	50	60
Sedimentary rock or coal	Minimum	30	35	45	45	50	60	55	60	70
	Average	55	60	70	70	75	85	80	85	95
	Maximum	70	75	85	85	95	100	95	100	110
Timbered (5-ft centers)	Minimum	80	85	95	95	100	110	105	110	120
	Average	95	100	110	110	115	125	120	125	135
	Maximum	105	110	120	120	125	135	130	135	145
Igneous rock	Minimum	90	95	105	105	110	120	115	120	130
	Average	145	150	160	160	165	175	170	175	195
	Maximum	195	200	210	210	215	225	220	225	235

Table 2. Size and characteristics of airborne solids (compiled by W. G. Frank)*

[illegible]

*It is assumed that the particles are of uniform ...

coal dust. The common physiological diseases are mostly various pneumoconioses. Preventive measures are to suppress dust at the source with water sprays, foam, fog, wetting agents, or dust collectors. Additional protection may be provided through suitable respirators. Accumulations of explosive dust should be removed and inert material, such as rock dust, applied to surface areas. The dust hazard can be determined by systematic sampling of airborne ...

centrations of the sample can then be determined by count using a microscope or microprojector. See DUST AND MIST COLLECTION.

Air temperature and humidity. Temperature rise in mine workings is from (1) heat conducted from surrounding strata because of the thermal gradient (depth per °F rise in temperature), (2) adiabatic compression of descending air columns (approximately 5.5°F/1000 ft); and (3) heat from oxidation of minerals.

The factors that influence humidity are: rise in dry-bulb temperature; volume changes caused by pressure and tempe ature changes; and moisture picked up from shafts and roadways.

The air temperature approximates the temperature of adjacent walls; seasonal temperature changes are noticeable only short distances underground. Workers' efficiency (Fig. 3) is dependent upon temperature, humidity, and motion velocity of air. A solution to excessive air temperatures is air conditioning by passing the air currents through heat exchangers filled with chilled water.

Natural ventilation. Natural ventilation is induced by differences of total weights of air columns for the same vertical distance. Natural draft may operate with or against the mechanical draft (Fig. 4) or may be the only source of pressure. Natural draft pressures at standard density can be estimated as 0.03 in. water-gage for each 10°F average temperature difference per 100 ft increment vertical elevation. For accurate determinations the average air densities of influencing air columns must be calculated.

The formula for determining air density is

$$d = \frac{1.327}{460 + T}(B - 0.378VP)$$

where d is the density in lb/ft³, T is the dry-bulb temperature, B is the barometric pressure, in inches of mercury, and VP is the vapor pressure of water at the dew point, in inches of mercury.

With most calculations, the vapor pressure influence can be ignored. The simplified formula then is

$$d = \frac{1.327}{460 + T}(B)$$

Auxiliary ventilation. This term applies to booster fans and auxiliary fans. A booster fan is placed underground and operated in series with the main fan to increase ventilating pressure of one or more splits of the ventilating current. The booster fan in effect reduces the mine resistance, thereby increasing the air quantity circulated. An auxiliary fan is a small fan installed in the air current to divert, through air tubing or ducts, a part of the ventilating current to ventilate some particular place or places. In metal mines, they are used to ventilate developing drifts, raises, crosscuts, and stopes. In

Fig 3. Graph of influence of humidity and temperature on worker efficiency.

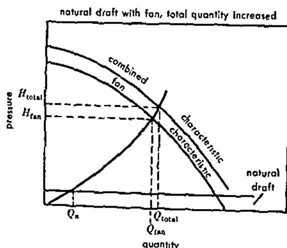


Fig 4. Graphs of influence of natural draft to mine ventilation.

coal mines, they are used to conduct air to working faces.

[D.S.K.]

Bibliography: G. E. McElroy, *Engineering Factors in the Ventilation of Metal Mines,* U.S. Bur. Mines Bull. 385, 1935; R. Peele and J. A. Church (eds.), *Mining Engineer's Handbook,* 3d ed., 2 vols., 1941

MINE ILLUMINATION

Mobile illumination. Cap lamps, flashlights, portable hand lamps, trip lamps for haulage cars, and lights in mobile machinery provide mobile lighting. The carbide lamp has been outmoded by the safer and more efficient electric cap lamp operated by a 4- or 5-volt battery. Such a lamp, using a polished reflector, can produce a beam candlepower of 1000 candles in a small spot. Less light with greater spread is obtained with a matte reflector.

Permissible continuous mining machines, shuttle cars, and other mining production equipment are fitted with explosion-proof 150-watt, 120-volt lights, front and rear. All permissible lights for use in gassy or dusty mines in the [illegible] lights at intervals of 20–90 ft. Incandescent types, 50–150 watts, dc operated, may be connected singly or in groups in ventilated haulageways but not in gassy areas. Fluorescent lighting, first approved in 1957 for gassy or dusty mines, employs 14-watt lamps on ac power (Fig 5).



Fig. 5. An installation of permissible fluorescent lights in a coal mine. (*U.S. Bureau of Mines*)

Fluorescent lights have proved very effective in gassy areas, along beltways and shuttle-car roadways, and have been credited with reduction in accidents, better morale among the men, and better production records. Three-wire conventional grounding or two-wire isolated ungrounded systems [illegible] facili[illegible] The [illegible] p for underground photographic use in 1957. [R.S.I.]

MINE DRAINAGE

It is necessary to prevent water infiltration into mines from the surface or from sources underground. Control, transportation, and ultimate disposal of such waters constitute mine drainage.

Mine drainage varies in facilities and importance, at different localities, according to conditions of the mine water. There must be a careful study of its occurrence, corrosive and erosive character, volume to be handled, and the handling and disposal facilities that can economically provide efficient drainage. Major facilities consist of flumes, dikes, and storage ponds to prevent or minimize seepage of surface waters, underground ditches and sumps to collect, store, and control mine water, and pumping plants and drainage tunnels for the disposal of the mine water.

Control of stream pollution by mine drainage has become highly important in engineering for aesthetic and economic reasons.

Sources and types of mine waters. Physical characteristics of the individual mining areas create distinctive sources and types of mine waters. Water enters directly into open-pit or underground workings, seeps through pervious strata, drains from normal water channels both on the surface and underground, and infiltrates from impounded waters into mine workings through cracked, crushed, and broken formations.

Mine waters vary widely in character in different mines because of the different geological formations. Waters from mines may be classed as alkaline (pH 7–14), neutral (pH 7), and acidic (pH 0–7). They may carry abrasive matter in suspension and be erosive to a degree dependent on the amount of abrasive substances. Although ground waters are normally alkaline, analyses from the majority of mines show considerable variation because the waters contain varying quantities of dissolved sulfate salts. Acidic water formed by such salts in the majority of mines has resulted in tremendous losses by corrosion, especially to pipes and pumping equipment.

Sumps, drains, and tunnels. The infiltration of water into mines, the volume of water to be handled, and pumping facilities vary with the seasons The storage of mine water until it can be disposed of satisfactorily is of great importance. Ample sumpage where mine water can be stored is necessary in any drainage system. Sumps should be provided near pumping stations, have ample capacity, and be arranged so as to permit easy cleaning. If practicable, they should be designed to provide gravity feed to the pumps.

Drains, flumes, or ditches. These are used on the surface and underground to divert and convey water and thereby to prevent stream pollution and inflow through crop caves, strippings, cracks and fissures, and stream beds. Large flumes are used for those purposes in subsided areas, particularly where flash floods or quick runoff occur. Because ditches are usually made in mine haulage roads, they should be designed to carry the maximum drainage.

Drainage tunnels between different working levels are preferred means of handling and keeping mine water from reaching lower mine workings. Drainage tunnels from a mine or a group of mines to some disposal point situated favorably with respect to surface disposal areas are a means of draining mines economically and saving reserves of minerals that would be lost otherwise. They are advantageous over a long economic life of a mine for handling large quantities of water that sometimes could not be pumped to the surface. Although the initial cost is high, their upkeep is comparatively low.

Pumping. As long as mines are operated, mine drainage will most likely be done in whole or in part by pumps. Because mine water is usually acidic, suitable metals or alloys should be used for mine pumping equipment. The designing engineer should consult with qualified technicians before attempting to solve corrosion problems Centrifugal pumps are both horizontal and vertical and known as standard, deepwell, and shaft. Displacement pumps, piston and plunger, are being discarded. Mine tailing pumping is sometimes done by means of centrifugal sand pumps. Deepwell pumps in a shaft or bore hole have proved successful for emergency use and for unwatering abandoned mines. Pumping systems and controls must be designed to handle the maximum inflow of water in mine workings. See PUMP. [S.H.A.]

Bibliography: American Standards Association, *American Standard Recommended Practice for Drainage of Coal Mines*, U.S. Bur. Mines Bull. 570, 1957; S H. Ash et al., *Corrosive and Erosive Effects of Acid Mine Waters on Metals and Alloys for Mine Pumping Equipment and Drainage Facilities*, U.S Bur Mines Bull. 555, 1955; S. H. Ash, H A Dierks, and P. S. Miller, *Mine Flood Prevention and Control, Anthracite Region of Pennsylvania, Final Report of the Anthracite Flood-Prevention Project Engineers*, U.S. Bur. Mines Bull. 562, 1957.

STORAGE AND SHIPPING

Although a task of great proportions, at least with respect to coal mining, the operation does not require extensive mine operating facilities. Coal is mined only when market demands for the product have been established and shipment to the consumer is made immediately. Storage facilities at the coal mines seldom provide more than surge capacity and are used to permit continuous operation of mine during temporary shutdown of preparation plant or loading outlets. Where storage of coal is required to provide a continuing fuel supply, necessary storage facilities are provided by the purchaser at the point of consumption.

Ore storage facilities are needed at mines because market cycles and changing seasons prevent steady ore shipments. At mines with mills or washing plants, different operating rates require surge capacity to assure uniform flow of ore to each unit.

Bulky ores, such as those of iron, limestone, gypsum, sulfur, sand, and gravel, are kept in stockpiles on the ground Iron mines, for example, store 5,000,000–6,000,000 tons of ore each winter in stockpiles Common practice is to build the piles about 50 ft high by dumping from a trestle or aerial tramway. At open-pit mines, stockpiles are formed with dump trucks or stacker conveyor-belts. Ore is withdrawn from stockpiles with mechanical loaders and occasionally through tunnel draw holes.

Underground mines, in themselves, provide a certain amount of ore storage in the stopes and in skip pockets, thus providing uniform flow to mills. At mills and ore-treatment plants, surge bins, with capacities of 50–3000 tons for coarse ore, are used ahead of primary crushers and bins of 5- to 50-ton capacity ahead of fine crushers and screens. Ore concentrate storage facilities range from 10-ton capacity at small mines to a 1,700,000-ton capacity unit, at Silver Bay, Minn., for storing iron ore pellets during the winter months.

Bulky ores are shipped by truck, in open railroad cars, and in bulk-cargo vessels. Concentrate is shipped short distances uncovered in standard dump trucks, and long distances by rail in closed box cars. Most mineral concentrates are shipped bulk cargo in ocean-going vessels.

Mine refuse and tailings disposal. Alm of the product received by coal preparati

is mine refuse and tailings. Depending upon the terrain and the availability of disposal areas to the mining operation, refuse is hauled from less than 100 ft to more than 2 miles to the disposal site. Trucks, carryall scrapers, conveyors, aerial trams, and hydraulic systems are the most popular means of transporting mine refuse away from the coal preparation plant.

In the ore industry, whenever possible, mine and mill waste is stored so that valuable mineral constituents can be reclaimed, or it is stocked for possible later use. Coarse tailing is loaded into cars or trucks and stockpiled. Fine tailing is impounded by dams made from the waste material. These dams are formed of the sand portion of tailing slurry transported to the site in launders, or running-water transporters, arranged to discharge the slime to the pond and the sand to the dam. Flocculants are often added to ensure clear-water overflow. At some underground mines, tailing is deslimed, dewatered, and returned to the mine for ground support.

Sorting, washing, and screening plants. About 60% of the coal produced in the United States is subjected to some form of mechanical cleaning. These plants use a variety of mechanical cleaning methods ranging from rudimentary equipment to multimillion dollar installations operating with a high degree of efficiency. Virtually all coal mines provide means for sizing the raw coal loaded from the mine into the size grades best adapted to the

coal washing facilities are used, the large-size lumps are frequently subjected to visual inspection and the refuse material discarded. Only in isolated cases can hand-picking be justified economically on coal less than 2 in. in size. Early preparation plants used stationary bar screens over which the coal

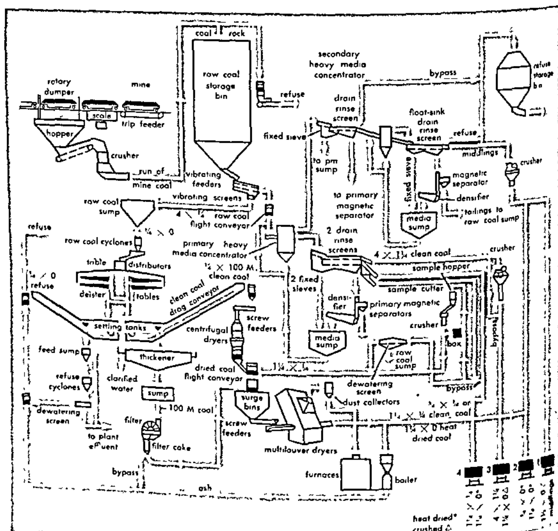


Fig 6 Flow diagram of a complex coal preparation plant. Dimensions in inches unless designated as microns, μ. (Link-Belt Company)

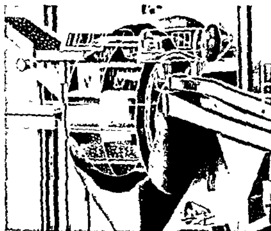


Fig. 7. Mechanical sorting and washing of coal. (Link-Belt Company)

flowed by gravity. The fine material passing through the bars was discarded as waste. With the development of coal combustion units capable of burning small-size coal, shaker and vibrating screens became commonplace in most coal preparation plants

Coal and mine tipples usually house the washing, screening, and loading facilities of the mine The early tipples were disproportionately large compared to capacity, but the modern tipple is a masterpiece of efficient design embodying automation to a considerable degree without retarding the ability to supply a satisfactory product covering a range of quality specifications. [W.L.C.; H.T.R.]

FIELD SAMPLING AND ORE ESTIMATION

These processes and activities are somewhat the field counterparts of laboratory assaying. However, the samples so derived may be sent to the laboratory for more refined analysis (*see* ASSAYING) Applications of statistical mathematics and theory of probabilities to sampling are discussed elsewhere (*see* STATISTICS).

Sampling. Ore sampling is the process of taking a small portion of ore in such a manner that the portion will be *representative of the whole* in respect



Fig. 8. Compact, efficiently designed coal tipple containing automatic machinery for a processing flow similar to that of Fig. 7. (Link-Belt Company)

to some quality or characteristic. The quality with which sampling is concerned in mineral deposits is usually the chemical analysis or grade of the material for commercial use. This is also called assay value When a mineral deposit contains sufficient valuable ingredients to be mineable at a profit over costs, the mineral-bearing material is called ore, and the deposit an ore body. It is the assay value of the ore body which is of interest.

The unit of valuable material varies among ores; for example, for the precious metals (gold, silver, platinum), the assay is given in troy ounces of metal per short ton of ore, for the nonferrous metals (copper, lead, zinc, tin), the percentage of metal in the ore is used, and for other metals (iron, manganese, chromium), a more complete chemical analysis is used which may include the percentage or ratio of other specified ingredients in the ore.

Mineral deposits are seldom exposed to view to any appreciable extent. Outcrops may appear at the surface, and if the deposit is being exploited, there will be limited underground exposures in shafts and other passageways. In addition, drill holes are used to investigate hidden deposits.

Three widely used methods of sampling are referred to as channel, chip, and bulk sampling. They are essentially the same in principle but differ in field procedure.

Channel sampling. In this sampling method, grooves are cut or chiseled across the exposed face at specified intervals (Fig. 9*a*) The grooves are about 4 in. wide by 1 in. deep and are preferably cut at right angles to the inclination of the vein or formation. Material cut from the channel is the sample It is carefully saved for chemical analysis. If there is too much material from the cut, it is mixed thoroughly on the canvas on which it is caught and quartered down to a suitable amount. Long channels (across a thick deposit or lengthwise of a passageway) are often cut in 5- or 10-ft sections, each section representing a sample.

used in testing broken ore, such as ore in transit, is essentially the same in principle as chip sampling.

Bulk sampling. A substantial amount (perhaps several tons) of material is taken at each test location. This bulk is then either crushed and sampled or used for a mill test.

Sample calculations. A mineral deposit is 3-

known.

To determine the average assay value of the whole deposit or some designated part of it, it is necessary to combine the values of all sampl located within the selected limits. To accompl this, the several samples must be related to oth ording to the portion of the deposit

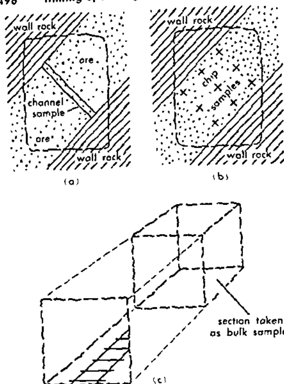


Fig 9. Sketch diagrams within a mine drift of major field sampling processes. (a) Channel sampling. (b) Chip sampling. (c) Bulk sampling.

each represents and correspondingly weighted for mathematical resolution to an average assay value. In basic terms, each sample is considered to represent the volume of material (ore) surrounding it and extending halfway to each adjacent sample. This volume, or a proportional factor, is the true numerical weight of the sample assay value when combining it with other assays.

When sample assays are used to determine the over-all (average) grade of a mining face (either to guide progress of work or preliminary to estimating available ore tonnages), each assay is weighted according to its area of influence (assumed to extend halfway to adjacent samples) because the third dimension (depth or thickness) is considered identical for all points on the face. This procedure does not violate the principle of volume weighting previously noted. Face averages may then be combined on a volume basis to determine the average assay of the related body of ore.

Figure 10 is a diagram of a channel-sampled face to illustrate the principle of weighting. Since only the average grade of the exposed face is desired, no third dimension is involved and since the channels are uniformly spaced, the interval between them (second dimension) may also be omitted pro-

sampling and calculation of average mineral content, estimation of ore for mining implies recoverable values and hence involves efficiency of mining (mining losses plus ore left in place for support), and milling and smelting losses. To convert volume of ore into tonnage requires testing of the deposit for specific gravity along lines similar to sampling for assay values. [RDP]

MINE EVALUATION

An appraisal of the monetary worth of a workable mineral deposit generally arrived at by experts in the mining industry. Such appraisal weighs the values of recoverable mineral content, operating efficiency when developed to production, and marketability of product. There are two parts to an evaluation: determination of factors, and calculation of present worth.

Determination of factors. Although past records may serve as a guide to the producing ability of the business and to indicate the cost-price relationship that may have obtained, the evaluation, as such, is concerned only with estimated future production and profits.

To estimate future operations the following features must be examined, analyzed, and verified: (1) grade and size of ore body and its remaining recoverable portion; (2) suitable production schedule, methods of mining and treatment, and corresponding plant and equipment requirements; (3) costs of production, treatment, and transports tion; (4) grade, market and selling price of product, and operating profits; (5) production life; and (6) interest rates suitable to the business risks involved.

Size, shape, attitude, and quality of the mineral deposit are determined by geologic study (and maps, if available), by inspection of the deposit and sampling of exposures, and by test drilling and sampling so planned as to reveal hidden areas. Quality is expressed as the amount of valuable constituent per commercial unit of material, such as

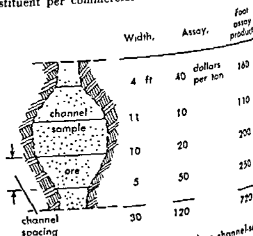


Fig 10. Diagram of weighting based on channel-sampled face of ore body. Computations from this case are Foot assay product ÷ sum of widths = average assay hence, 720 ÷ 30 = $24/ton

by troy ounces of gold per ton, or by percentage of copper, lead, or zinc, or by *full chemical analysis*. Structural quality and metallurgically undesirable constituents may also be included factors in the analysis. When the constituent is sufficiently valuable to warrant exploitation, the mineral-bearing material is spoken of as ore and the deposit is called an ore body.

Methods of *mining*, whether by *open-pit* or underground excavation will be determined by the depth, size, and shape of deposit, by the structural strengths of ore and surrounding rock formations plus considerations of output rate, and by the time and *investment required for development*, *plant*, and equipment.

Output rate, mechanization, and selection of specific method and practices from applicable alternatives are considered in estimating operating costs. *Production life is a function of recoverable* reserves and annual output.

In contemplating investment in a mining operation, consideration must be given to business and market risks which will be related to the industry, *location of the property*, *political environment*, taxation, problems of climate, transportation, labor force, and probable hazards of the specific operation.

The process of gathering and analyzing data is spoken of as the examination of the property. When complete data are needed, as in the case of an operating property or one with substantial considerations at stake, a formal, or thorough, examination is made. For less-developed properties a *preliminary* examination may suffice to show whether or not to go ahead with detailed study.

Once the examination of the property is completed and essential factors for evaluation are established, including suitable interest rates on capital to be *invested*, *the remainder of the evaluation* consists of mathematical calculations to reduce the estimated future earnings to present worth.

Calculation of present worth. The evaluator, in the second part of the evaluation, faces the unique feature of the mineral industry, namely *depletion*. Minerals are exhaustible resources. When mined, they cannot be replenished, except through eras of geologic time. Unlike the agricultural industry where production may be regrown next year, a mine can *produce the minerals from its ore body* only once; the producer must then move on to new ore bodies to continue operations. This feature, depletion, has long been realized as fundamental to evaluation. Courts have recognized the principle of depletion *in respect to mineral resources* (*State of* Minnesota, Supreme Court 30723, 1935). Income tax authorities also have assented by allowing percentage depletion as a deduction in calculating taxable profit.

In view of *depletion*, *the purchaser of a mineral* deposit or a mining business must recover his capital during the remaining productive life. In evaluating *mineral* property, capital recovery is provided for in the mathematical calculation of present worth of estimated future profits. This is done by separating annual net profit into two parts, annual return of capital, and annual return *on* capital. The annual *increments of capital return are considered* as sinking fund payments which, when accrued at safe interest, will amount to the total invested capital at the end of the productive life.

The remainder of the annual profit is taken by *the investor as interest on his capital investment*. By this procedure the investor receives interest on his capital during the productive life and at the depletion of the ore body, has intact his full original capital with which to start a new operation. *This type of "redemption fund" calculation follows* the Hoskold premise for mine valuation.

The Hoskold equation for determining the present value of anticipated uniform annual income may be expressed as

$$V_p = \frac{A}{[r/(R^n - 1)] + r'}$$

where V_p is the present value, A is the annual income (*profit*), r *is the safe interest rate* (*for sinking* fund), r' is the rate of interest on invested capital, n is the years of productive life, and $R = (1 + r)$ or $1.00 plus one year's interest.

When estimated future income results in nonuniform *annual profits*, *a variation of the above equa*tion, embodying the same basic principles, may be used to calculate present value.

Capital outlays for plant, development, and taxes prior to production are deducted from present value *of earnings in arriving at purchase or sale valua*tion. Interest on working capital and taxes are expense items to the operator.

Royalty is a fee paid to the owner for the privilege of working a *mineral* deposit. Although gener*ally viewed as division of gross profit between fee* owner and operator, royalty represents gross profit to the fee owner, or lessor, and a cost of production to the operator, or lessee. [R.D.P.]

Bibliography: J. A. Grimes and W. H. Craigue, *Principles of Valuation*, 1928; R. D. Parks, *Examination and Valuation of Mineral Property*, 4th ed., 1957.

Mining safety

The advancement of mine safety in the United States began with establishment of the Federal Bureau of Mines on July 1, 1910. The average of more than 2500 deaths in coal mines annually during the preceding 5 years, including an average of almost 500 deaths annually due to gas- and dust-explosion disasters, provided the impetus for the Bureau's enabling act. Previously some attempts were made to improve safety *in mines through state mining* laws, such as the Pennsylvania Anthracite Law enacted in 1891. However, virtually all comprehensive state mine safety laws were enacted after the Federal Bureau of Mines was established.

The best results in mine safety are achieved through accident-prevention education and continu-

ing cooperation among management, labor, state and Federal agencies, and independent associations affiliated with the mining industry.

Mine hazards and disasters. Regardless of the kind of mineral produced, the principal mining hazard is fall of roof or ground, the cause of approximately half of the annual fatalities and about one-fourth of the serious nonfatal injuries since records have been kept. Transportation accidents rank next in seriousness. Formerly explosions ranked third as the cause of fatal injuries, but this situation no longer prevails. With the advent of permissible explosives and blasting devices, under Bureau regulations, accidents from explosives have become relatively rare.

The rapid rise in mine mechanization has increased the number of injuries from electricity and electric-powered machinery. Common causes of injury are handling materials and falling, but except for falls in shafts, these accidents seldom result in fatalities.

Explosions. Although the incidence of mine explosions has diminished considerably, they have not been eliminated. For example, in the year 1957, major gas explosions in two bituminous-coal mines caused 37 and 11 deaths, respectively, and 20 other gas ignitions caused a total of 7 deaths.

In the early days of the Bureau frequent explosion disasters kept its personnel active recovering the dead, rescuing the survivors, and restoring the mines to normal. Permissible mine rescue apparatus was developed and approved by the Bureau to facilitate rescue and recovery operations, and the Bureau developed and presented basic and advanced mine rescue ...

t ...

workmen, especially in isolated districts, would be qualified to take care of the injured until medical assistance could be obtained. Approximately 2,000,000 persons have completed the Bureau of Mines first-aid course and 110,000 the mine rescue courses.

Many explosions during the first quarter of this century were caused by black blasting powder—a long-flame explosive that frequently ignited gas or

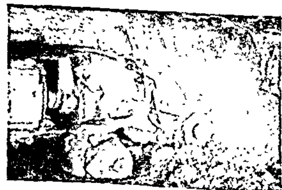


Fig. 1. Cutting head of ripper-type continuous miner with water sprays for suppressing dust at the coal face (U.S. Bureau of Mines)





Fig. 2. (a) Dry rock-dust application. (b) Wet (slurry) rock-dust application; note lack of "dustiness" (U.S. Bureau of Mines)

coal dust or both. Studies and investigations by the Bureau's Explosives Section led to development and manufacture of permissible explosives and blasting devices. Use of these, coupled with elimination of black blasting powder by law in all but the very small coal mines, has greatly diminished the incidence of disasters initiated by explosives.

Most widespread explosion disasters in bituminous-coal mines were initiated by ignition of a body of gas (methane) and propagated by coal dust throughout the mine workings. Tests and experiments by the Bureau of Mines with dusts from coal beds throughout the United States and elsewhere proved conclusively that the dusts from coals having a volatile-combustible ratio

$$\text{Volatile-combustible ratio} = \frac{\text{volatile matter}}{\text{volatile matter} + \text{fixed carbon}}$$

over 0.12 are explosive and may propagate an explosion. The higher the volatile-combustible ratio ... will explode. Virtually ... (not anthra... therefore ex...

plosive. Although water sprays are advocated to allay dust at its source in cutting, drilling, blasting and loading operations, it is impossible to suppress all dust. Some of the finer, more dangerous coal dust will become airborne, and as little as 0.05 ounce of it suspended in 1 cubic foot of air space may propagate a dust explosion. The only sure method of preventing propagation of an explosion by coal dust is adequate rock dusting. To render coal dust nonexplosive an inert substance, such as finely pulverized limestone dust, should be applied

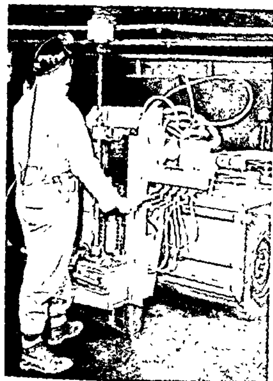


Fig. 3. Rotary drill with dust collector approved by Bureau of Mines for vertically upward drilling. (U.S. Bureau of Mines)

on the roof, ribs, and floor to within 40 ft of the faces (including the last open crosscut) so that a minimum of 65% incombustible matter will be maintained at all times. The incombustible minimum should be increased 1% for each 0.1% methane in the ventilating current. This rock-dusting standard should be adhered to in all bituminous-coal and lignite mines except those that are very wet.

Pure dust explosions are rare compared with those caused by gas ignitions. Dust explosions are caused mainly by nonpermissible explosives, nonpermissible use of explosives, or electric arcing. Proper rock dusting will prevent the spread of dust explosions.

Gas ignitions are triggered mainly by electrical sources or by frictional sparks of cutting bits. To produce combustion—slow in a fire or rapid in an explosion—three elements are required: combustible material, oxygen, and a source of ignition. Remove any element and combustion is impossible.

Fires. Most mine fires result from electric arcing or excessive heating in intimate contact with combustible material, such as coal, timber, flammable oil, or conveyor belting. Many fires are the aftermath of explosions; some have initiated explosions. Roof falls, bringing down trolley wire and causing short circuiting and arcing which set fire to timber or coal, are a common cause.

The frictional heat developed by "freezing" of pulleys, causing conveyor belts to ignite, has been the source of many mine fires. However, fire-resistant conveyor belts, constructed according to Bureau of Mines specifications, have reduced the incidence of fires from this cause. The Bureau tests conveyor belts and approves those that meet its requirements for resistance to fire and damage.

Trailing cables are a principal factor in mine fires, mainly because of poor splicing and overheating of combustible insulation. With Bureau encouragement thermoplastic, fire-resistant trailing cables are available to minimize this hazard.

Fire-fighting equipment should be installed at

extinguishers should be selected with care: water or aqueous solutions should not be used on electrical or oil fires; rock dust or sand is acceptable for this purpose. Extinguishers containing carbon tetrachloride or other chemicals, that decompose in contact with heat to form lethal gases, are dangerous in confined areas.

A great danger in fighting mine fires directly is disintegration of the roof by heat, followed by cave-

noxious gases.

If a fire cannot be fought directly, it should be brought under control by sealing, which depletes the oxygen in the atmosphere so that it cannot support combustion, and the fire is extinguished. Sealing is especially dangerous in a fire area in which gas (methane) is liberated, as it involves a race between the sealers to complete their job and get out of the mine before the lower explosive limit is reached when the fire could ignite the gas and initiate an explosion.

Provision should be made to permit sampling and analyzing the atmosphere behind fire seals. Premature opening of fire seals must be guarded against or restored ventilation may cause the fire to break out again and possibly set off an explosion. In general, analysis of the sealed atmosphere should show no carbon monoxide and very low oxygen content, preferably less than 1%.

Flooding a fire area should be the last resort because of the danger of impounded water, damage to mine workings, and the possibility of damming hot gases in high places, which may cause the fire to break out again when the water is removed.

Flooding. Flooding or disastrous inrushes of water or "running" ground generally have been more prevalent in metal than in coal mines. The Pennsylvania anthracite mines have been especially affected by flooding, but as many of them have steeply pitching seams, the mining methods resemble those at many metal mines. Mining below "buried valleys" containing unstable, or "running," ground, with attendant loss of life, was a fairly common occurrence in the Pennsylvania anthracite region, but thorough exploration by drilling and accurate mapping have reduced the incidence.

Impounded water in abandoned areas of any ... hazard to workmen in lower-lying

200 ft of abandoned workings containing impounded water, the extent of which cannot be accurately determined, advance boreholes at least 20 ft deep should be drilled, close enough to each other to insure that the place will not cut into the inundated area unexpectedly. When the drillings of such advance holes become damp or gas is encountered, mining should be discontinued, and the holes should be plugged immediately with prepared soft wooden plugs. These will swell when wet and wedge tightly in the borehole.

When impounded water is tapped in a mine the work should be done only by closely supervised, experienced workmen, and all others should be barred from the mine. Flame safety lamps and other atmosphere-testing devices should be available. Only closed lights should be used for illumination.

Other sources of flooding are surface bodies of water. Their locations should be plotted accurately on the mine map. Flooding may occur because of excessive precipitation, causing swollen streams to overflow. If inclined or vertical mine openings are in a flooded area, sudden inrushes may occur. The safeguard against flash floods is to have all mine portals well above any known flood stage.

Gases and ventilation. Mine gases may be classed as follows.

1. Gases liberated during normal mining operations, such as methane in coal, limestone, and salt mines; radon in uranium-ore mines; and hydrogen sulfide in salt and gypsum mines. As blasting is considered a part of normal mining operation, the smoke and gases generated by explosives are included in this class.

2. A wide range of gases or combinations of gases liberated or generated by abnormal conditions, such as mine fires and explosions. The most common and dangerous are carbon monoxide, oxides of nitrogen, and sulfur dioxide.

In normal mining operations, including blasting, good ventilation (properly directed by auxiliaries such as stoppings, overcasts, curtains, and fire-resistant brattices) is necessary to dilute, render harmless, and carry away the gases as they are liberated.

Gas-detecting devices, including flame safety lamps, are available to detect any kind of gas that might be evolved during normal mining. The flame safety lamp is the best known and is useful to detect methane, as well as oxygen deficiency.

A serious problem has developed with modern continuous-mining operations in face areas of gassy coal mines. Some continuous miners are so long and fill the working area to such an extent that it is impossible to extend a line brattice close enough to the face to sweep away the methane as it is liberated. Frequently a continuous miner discharges coal to the mine floor, where it is picked up by a loading machine. With the floor acting as a "surge bin," accumulated coal tends to form a "pocket" at the face, increasing the difficulty of ventilation.

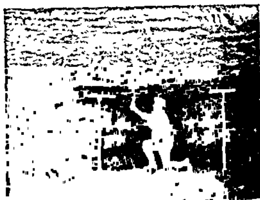


Fig. 4. Foreman testing for gas (methane) with flame safety lamp at face of trackless, roof-bolted entry before rock-dust application. (U.S. Bureau of Mines)

Although safety (testing) lamps usually are available in face areas, some serious gas ignitions which were initiated by electrical machinery and accompanied by loss of life, have resulted because the gas was not detected or, if so, nothing was done in time to prevent ignition. The Bureau of Mines is sponsoring research and development of an automatic methane detector, installed as an integral part of each electric-powered machine operating in face areas, that will shut off the power when the methane reaches a predetermined concentration, well below the danger point. Separate methane alarms are now available, but their warnings can be dangerously ignored by workmen.

Cave-in of excavations. Falls of roof and ground are the principal causes of fatal and serious nonfatal injuries. They can be prevented by establishing and enforcing a system of minimum artificial support. Roof, back, or ribs should be frequently and thoroughly examined and tested to determine whether the established minimum of roof support should be augmented. Only thorough study and experience can determine the adequacy of a roof support system for a particular operation.

Conventional roof support consists of various forms of wood or metal (usually steel), such as props, crossbars, cribs, arches, and various combinations thereof. An ingenious method known as roof bolting, or rock bolting, is used extensively in both coal and noncoal mines; for further details see MINING, UNDERGROUND. Roof or rock bolting is not universally applicable and sometimes must be used in conjunction with conventional timbering. Where it can be adopted, it has definite advantages over conventional timbering, because vertical supports that offer resistance to ventilating currents or that can be dislodged by portable equipment are eliminated, thus facilitating mechanized mining. Roof bolting is generally unsatisfactory in roof that has sagged. If roof bolts are installed before

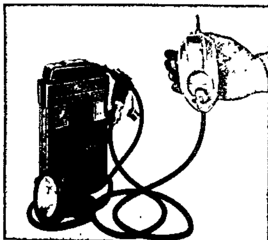


Fig 5. Mine Safety Appliances Company's E-2 methane detector, attached to electric cap-lamp battery. (U S. Bureau of Mines)

roof deformation, they will consolidate the strata into an adequately strong beam, provided suitable anchorage can be obtained.

The Bureau of Mines is experimenting with injection of quick-setting liquids, such as epoxy and polyester resins, under high pressure, through boreholes into stratified mine roof. When the liquid solidifies, the roof strata are actually glued together. Test results as yet are inconclusive but indicate considerable promise for the method, alone or in combination with roof bolting.

Conventional timbering or roof bolting serves as support for secondary roof. Primary roof is supported by pillars. When pillars are too small to support massive overlying strata, large cave-ins will result. Depending on the characteristics of the strata, failure of pillars will result in squeezes or creeps; but violent outbursts or "bumps" sometimes will be caused by sudden collapse of pillars under heavy rock cover. Obviously the collapse of pillars (sudden or gradual) is a serious working hazard. Sometimes it can be avoided by judicious removal of pillars or release of pressure by large auger holes drilled through affected pillars. This work is extremely hazardous and should be undertaken only by the most experienced workmen.

Mining diseases and health. Much fundamental research on health matters has been carried on by the Bureau of Mines, not only in laboratories but also in the field, in mines, mills, and industrial plants of the mineral and allied industries. These studies included the occurrence and prevention of dust diseases; the action of particulate atmospheric contaminants (dusts); the occurrence, properties, and toxicity of gaseous atmospheric contaminants —their effects, methods of recognition and alleviation of their dangers; health and sanitation in mines and mining communities; and mine ventilation To promote use of safe and effective dust-collecting devices in connection with rock drilling in coal mines, investigations are conducted on such equipment; and approvals are issued for those devices that meet the requirements of a prescribed schedule. As almost every type of respiratory hazard occurs in the mining industry, the Bureau of Mines has been instrumental in developing and maintaining standards for protective equipment. Manufacturers of respiratory-protective equipment generally seek to obtain Bureau approval for all new equipment. As a consequence, there are now available tested and approved self-contained breathing apparatus, gas masks, supplied-air respirators (hose masks, air-line respirators, and abrasive blasting helmets, hoods, or masks), and dispersoid (dust, fume, and mist) respirators.

Coal and free silica, or quartz (SiO_2), dusts now occupy the greatest attention of those interested in the health of miners. A limited statistical study conducted by the Public Health Service for a 5-year period (1950–1954) revealed that, out of a total of 10,152 cases of silicosis, there were 1637 cases (16.1%) in the metal-mining industry.

As a result of these findings, the Bureau of Mines in cooperation with the Public Health Service launched a study of environmental health hazards in metal mines. The studies will embrace (1) evaluation of the silicosis problem, (2) determination of dust concentrations and existing dust-control measures, and (3) investigation of other factors related to environmental exposure, such as free silica in airborne dust and size of dust particles. [M.J.A.]

Minitrack

A method of observing and accurately tracking missiles, satellites, and space vehicles by radio. The Minitrack (minimum-weight tracking) tech-



Fig 1. Location of Minitrack stations set up for International Geophysical Year satellite program.

N

north fine

minitrack electronics trailer

coarse and medium

west fine

east fine

calibration shelter

south fine

telemetering antenna

headquarters building

Fig 2. Typical station layout with coarse, medium, and fine antennas in N-S plane, and medium and coarse antennas in the E-W plane

nique consists of obtaining the angular positions of a satellite or vehicle carrying a beacon transmitter, through measurement of the phase difference between the signals arriving at two spaced ground antennas with fixed beams, as the vehicle passes across the beams. It is the radio equivalent of the double slit in optics but with the added advantage that the phase anywhere in the antenna lobe can be measured.

In connection with the International Geophysical Year satellite program, a Minitrack system was set up with stations spaced several hundred miles apart along a N-S line to form a radio "fence" (Fig. 1). In this way at least one station can intercept a satellite each time it circles the earth.

A two-axis Mi...

[illegible]

measurement, widely spaced antennas (500 ft or about 50 wavelengths for a 108-Mc signal) are used; hence there are many lobes or fringes. To determine in which of the fine lobes the signal is arriving, a second system of coarse lobes produced by closely spaced antennas is used. For satellite tracking, each antenna has a fan-shaped beam pattern oriented with its broad dimension along a N-S line. See SATELLITE, ARTIFICIAL.

[J.P.H.]

Bibliography: R. L. Easton, Radio tracking of the Earth satellite, *QST*, 40(7):38–41, 1956; J. T. Mengel, Tracking the Earth satellite, and data transmission, by radio, *Proc. IRE*, 44:755–760, 1956.

Mink

A carnivore, *Mustela vison*, of the family Mustelidae, famed for its great value as a fur animal. The mink is somewhat larger than similar weasels. The males are 20–24 in. long, the females somewhat smaller. It is rich brown in color, with a white chin.

Minks are semiaquatic animals, never found far from water, although individuals may wander greatly. Their principal food is the muskrat, and therefore minks serve as the principal natural con-



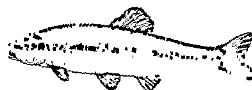
Mink, *Mustela vison*; length to 25½ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

trol of muskrats, which tend to develop overpopulations rather quickly. Minks are found throughout all of the United States and Canada except the arid Southwest and extreme North.

Mink farming is carried on profitably in many areas, and various special color mutations have brought excellent prices on the fur market. See CARNIVORA; WEASEL. [J.D.B.]

Minnow

A name technically limited to any member of the family Cyprinidae, but often used, although incorrectly, for any small fish. There are well over 1000 species in the Northern Hemisphere and in Africa, making this the largest of all fish families. Over 300 species occur in North and Central America. The minnows reach their greatest diversity in southeastern Asia. Most American forms are quite similar in spite of the great number of species. They are soft-finned with cycloid scales, toothless jaws, and with throat, or pharyngeal, teeth. There are rarely more than eight rays in the dorsal fin. Most minnows are brownish to olive on the back, with silvery sides. Several species develop reddish color on the breeding males in the spawning season, and a number develop tubercles, called pearl organs, on the head and pectoral fins during the breeding season.



Blunt-nosed minnow, *Hyborhynchus notatus;* length to 4 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

The carp and goldfish are Asiatic members of the family, but differ from other species in that they have a spine on the front of the dorsal and anal fins, and also have long dorsal fins. The largest American native minnow is the white salmon, *Ptychocheilus lucius*, a native of the Colorado River. Specimens weighing 10 lb are common, and white salmons weighing over 80 lb and growing more than 5 ft long have been reported. See CARP; CYPRINIFORMES; DACE; GOLDFISH; SHINER. [J.D.B.]

Miocene

The next to the youngest of the five major worldwide divisions (epochs) of the Tertiary Period (Cenozoic Era); the epoch of geologic time extending from the end of the Oligocene to the beginning of the Pliocene. See CENOZOIC; TERTIARY.

The term Miocene was originally proposed in 1833 by Sir Charles Lyell, who divided the Tertiary into Pliocene (youngest), Miocene, and Eocene. Subsequently, when it was realized that contemporaneous deposits in various parts of the world were being called upper Eocene by some geologists and lower Miocene by others, E. Beyrich, in 1854, introduced the term Oligocene for the time and beds in question between Eocene and Miocene.

PRE-CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

Tertiary					Quaternary	
Paleocene	Eocene	Oligocene	Miocene	Pliocene	Pleistocene	Recent

The Miocene Series includes all rocks formed during the Miocene Epoch, but the term is used most specifically with reference to the sedimentary rocks that were formed during this interval of geologic time. They contain the plant and animal remains that are the primary bases for identification of Miocene age.

Lithology. Miocene strata include all the common sedimentary types: marine, continental, and intermediate or transitional. They are typically unconsolidated to poorly consolidated and are present in many parts of the world. Particularly noteworthy are the Miocene strata of (1) the Atlantic and Gulf Coastal province of North America; (2) the Pacific border basins of western North America; (3) the North Sea area of northwestern Europe, the Mediterranean Sea area of southern Europe and northern Africa, and the intracontinental basins of central and eastern Europe and southern Asia; (4) the Siwalik region of the Himalaya Mountains; and (5) the coastal region of South Australia.

Igneous rocks of this age are best preserved in mountainous areas, terrestrial strata are best known in the continental interiors, and marine beds are most widespread along the continental margins in the coastal plains and continental shelves. Most Miocene strata are relatively flat-lying; many marine and marginal strata of this age occur at or below sea level but crustal disturbances have appreciably deformed some of the beds and have elevated them to considerable heights, as, for example, in the Alps and Apennines of Europe; the Atlas Mountains of northern Africa; the northern mountains of Chiapas, Mexico; the Pacific border region of North America; and the Himalaya Mountains of Asia. As many as 5000 meters of Miocene

strata are present in certain sedimentary basins, as for example in the Gulf Coastal area of Louisiana and Tabasco, Mexico, but in general the total thickness of these beds is appreciably less. Miocene sediments contain important quantities of oil and gas, ground water, clay, sand, marl, and other products.

Fauna. The Miocene is characterized by faunas (and floras) very similar to those of today. Among mammalian faunas, horses, cattle, camels, and elephants, distinguishable from their Recent counterparts only by details of tooth and bone structure, were the dominant forms. Gone were the large rhinoceroslike forms (titanotheres) that flourished in early Tertiary time. Perhaps the chief difference between Miocene terrestrial faunas and those of today lies in their distribution. Groups like the camels and elephants, now very much restricted geographically, were quite widespread in the Miocene. The marine fauna of the Miocene, with minor exceptions, has very nearly the same composition and distribution as that of today. *See* PALEOBOTANY; PALEONTOLOGY. [A.H CH.; C.E.M.]

Mira

The first star recognized to be a periodic variable. Called "The Wonderful" by the ancients, Mira was discovered by Fabricius in 1596 and has a period of 332 ±9 days. It is the prototype of long-period variables, and like most of its class, changes considerably from epoch to epoch. At times the maximum is as bright as second magnitude, and sometimes it is barely fifth magnitude.

Mira, o Ceti, has a faint companion discovered by R. G. Aitken after A. H. Joy predicted its presence from a study of the spectrum. The companion, at times, is about as bright as Mira is at minimum; the companion has a dwarf class B spectrum, with peculiar bright lines. The spectrum of Mira changes from M5e at maximum to M9e at minimum. The companion is also variable, which may account somewhat for the difference in depth of the minima of the system, which vary from about 8.5 to 9.5. *See* VARIABLE STAR. [M.W.M.]

Mirage

A name for a variety of unusual images of distant objects seen as a result of the bending of light rays in the atmosphere during abnormal vertical distribution of air density. If the air closer to the ground is much warmer than the air above, the rays are bent in such a way that they enter the observer's eyes along a line lower than the direct line of sight. The object is then seen below the horizon, the inferior mirage. If the air closer to the ground is much colder than the air above, the rays are bent in the opposite direction, arriving at the observer's eyes above the line of sight; the object then seems to be elevated or floating in the air, the superior mirage. More complicated or irregular stratification of air density causes apparent multiple reflection of the object and vertical or horizontal distortion of the image. Mirages can be seen most frequently along an overheated highway surface; the inferior mirage of the sky gives the impression of water reflection over a wet pavement, which disappears upon a closer viewing [A.S.]

Mirror optics

The science and technology of those optical surfaces called mirrors which, by means of reflecting rays of light, either revert optical bundles or focus them to form images. Mirror optics has always played an important role, particularly in astronomy, for many of the precision astronomical instruments employ reflecting elements. Since the mid 1930s, wider usage has been made of reflecting elements in optical systems because of the advantages they offer in producing high speed, wide angle imagery. Their more general usage has been made possible largely through the development of techniques for vacuum deposition of highly reflecting thin films on optical surfaces.

An optical surface which specularly reflects a large fraction of incident light according to the simple reflection law of geometrical optics is called a mirror surface. This law states that a specularly reflected ray of light lies in the plane containing the incident ray and the surface normal at the point of reflection, and that it and the incident ray form equal acute angles (θ in Fig. 1) with the normal on opposite sides (*see* OPTICS, GEOMETRICAL). A mirror surface is differentiated from a diffusely reflecting surface, which reflects light rays violating one or both of these conditions

Plane mirrors. These are used to redirect rays, for example, to reduce the over-all length of an optical system by "folding" the optical path, and to invert images in optical systems to compensate for inversions introduced by other optical elements.

A plane mirror forms an image of an object such that the mirror surface is perpendicular to and bisects the line joining all corresponding object image points. Object points lying in planes parallel to the mirror retain their relative positions in the image, while those lying in planes perpendicular to the mirror undergo a reversal in the image. Because this reversal occurs in only one of these planes, it is not possible to rotate an object into its mirror image.

The image is called a virtual image because light only appears to come from the image points. A real image is one in which light rays actually pass through the image points. *See* IMAGE, OPTICAL.

The location of the image depends only on the position of the object and the mirror surface, and not on the point from which the image may be viewed. When one looks at the reflected image of an object, he is, in effect, looking at the image through a transparent window the size and shape of the mirror. The portion of the image he can see is identical to that of a corresponding object viewed through the window.

Double mirrors. The virtual image of an object formed by reflection from two plane mirrors inclined at an angle θ to each other is found by ro-

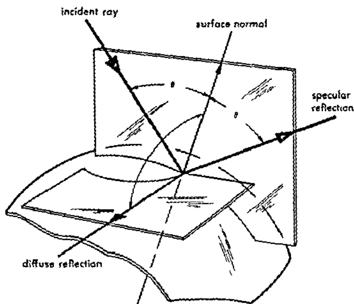


Fig. 1. Reflection.

tating the object through an angle 2θ about the line of intersection of the mirrors, the rotation being made toward the second mirror.

In this case, the position of the image depends only on the position of the object, the angle between the mirrors, and the orientation of their line of intersection. A rotation of the two mirrors about their line of intersection, keeping the angle between them constant, produces no displacement of the image.

An important use of double mirrors is to produce a given deviation of a ray of light, which remains unaltered regardless of any rotation of the mirrors about an axis parallel to their line of intersection.

A special case of particular interest occurs when the mirrors are at right angles to each other, that is, when $\theta = 90°$. In this case, the image is rotated through 180°, all points in a given plane perpendicular to the intersection of the mirrors being projected through the point of intersection an equal distance beyond the intersection. Any ray in this plane will, after two reflections, be oppositely directed and parallel to the incident ray.

Triple mirrors. A special case of interest, wherein reflections take place at three plane mirrs. occurs when the mirrors are mutually perpencular. This configuration is referred to as a iple mirror, corner cube reflector, or Amici cube. ɜe image is formed by projecting every object int through the vertex of the three mirrors a disnce beyond the vertex equal to that of the object int to the vertex. The image of a fixed object es not move during any rotation of the mirror sysm about its vertex.

If one views his own image in a corner cube reector, the eye through which the image is seen alays appears centered on the vertex, regardless of e orientation of the mirror. If the eyes are alterated, the vertex appears to jump from eye to eye. both eyes are used simultaneously, two vertexes re seen, one centered in each eye.

Another statement of this property of three mutually perpendicular mirrors to project the image through the vertex is that any ray of light incident on the mirrors will, after the three reflections, be oppositely directed, parallel to, but displaced from, the incident ray. It will lie in the plane containing the incident ray and the vertex, and the distance between the incident and reflected ray will be twice that of the vertex from the incident ray. Because of this retrodirective property, corner cube mirrors are called retro-reflectors.

The ability of two plane mirrors to convert an object space into an image space by a single rotation provides a simple basis for determining the location and orientation of an image formed by reflections from any even number of plane mirrors. The image formed by the first two mirrors is taken as the object of the next pair, and the image produced by these is located and serves as the object for the following mirrors. The image after reflection from $2m$ mirrors is determined by m rotations of space.

If only the orientation of the resultant image, not its actual location, is desired, all the rotation axes can be referred to a common center, and a single resultant axis and angle of rotation derived.

Prisms For many mirror applications, advantage is taken of total reflection occurring in prisms where the angle of incidence at a glass-air interface is greater than the critical angle. Such prisms can be made to provide both even and odd numbers of total reflecting surfaces, which are subject to the same treatment as ordinary mirror systems.

The right-angle prism (Fig. 2*a*) is used as a plane mirror to deviate a ray, usually through 90°. Since this provides a single reflection, reversing only one dimension of an image, for some applications the reflecting surface is broken into two halves at right angles to each other, the vertex being parallel to the original surface. This is called the Amici or roof prism, and it reverses both

dimensions in an image. The construction of the roof edge demands high accuracy in the angle between the faces and the sharpness of the edge.

The Porro prism is the same shape as the right-angle prism, but utilizes half of the hypotenuse face for the incident rays which, after reflection at the other two faces, emerge through the other half. Two of these prisms placed with their vertex edges at right angles to each other form the erecting system for binoculars to produce an erect image. *See* BINOCULARS.

The pentagonal prism, commonly termed Penta prism, provides two reflecting surfaces making a 45° angle, and therefore, produces a 90° deviation of a beam of light It is commonly used as an optical square to ensure that two lines of sight are accurately perpendicular. Since the rays are incident on the reflecting surfaces at an angle less than the critical angle, internal reflection does not occur, and these surfaces must be silvered.

The rhomboid prism is used to displace a beam parallel to itself. Like the Penta prism, any rotation about an axis parallel to the reflecting faces produces no deviation of the reflected beam.

The Dove or Delaborne prism provides a continuous rotation of a field through any deviation. As the prism is rotated about an axis parallel to the reflecting face and lying in a plane perpendicular to the refracting faces, the image is rotated through twice the angle.

The Pechan prism is a rotation prism similar to the Dove. It consists of two prisms with their hypotenuse faces separated by a plane parallel air gap. The incident beam is totally reflected at the first hypotenuse, but after reflection at the second surface (which is silvered), it is transmitted through the hypotenuse faces. It is then internally reflected at the emerging face and the hypotenuse, after an intermediate reflection at a silvered surface. It thereupon becomes parallel to its incident direction and passes through the emerging face. The Pechan prism improves over the Dove prism by providing incident and emergent faces perpendicular to the beam, and offering a very long optical path compared to the physical space it occupies.

All the prisms described are optically equivalent to a plane parallel plate of glass and introduce comparable aberrations into an optical system.

The advantages of employing prisms over simple mirrors lie in (1) the increased reflectivity provided by totally reflecting surfaces over mirrors (al-

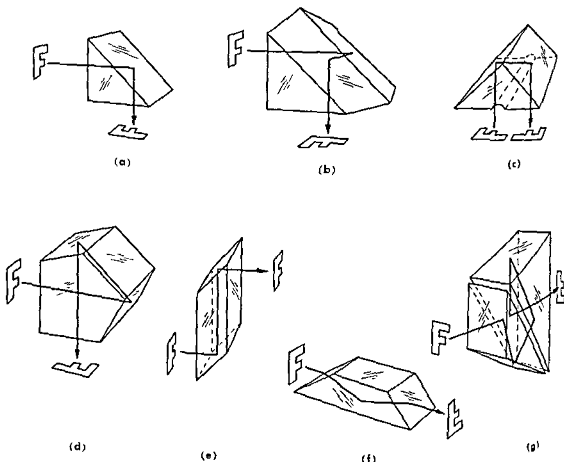


Fig 2. Prism types. (a) Right-angle prism. (b) Amici roof prism. (c) Porro prism. (d) Penta prism. (e) Rhomboid prism. (f) Dove prism. (g) Pechan prism.

though this may be partially offset by losses from surface reflections at the incident and emergent faces of the prism and by absorption in the transparent material from which the prism is made), (2) the stability with which angles can be maintained between the reflecting surfaces, and (3) the ease of ruggedly mounting the elements in an optical system without distorting the reflecting surface.

The disadvantages lie in (1) the increased cost of prisms over mirrors, (2) the optical aberrations introduced by utilizing a prism in an uncollimated beam of light (optically equivalent to placing in the beam a plane parallel slab of the same material as the prism and of a thickness equal to the total light path through the prism), (3) the bulk and weight of large prisms over equivalent mirrors, and (4) the difficulty in obtaining blanks of optical quality for large prisms. See ABERRATION, OPTICAL; PRISM, OPTICAL.

Beam-splitters. Plane mirrors are also used as beam-splitters when partial reflection and partial transmission of the energy in a light beam are desired. In many optical instruments, an image is viewed through a beam-splitter inclined at 45° to the beam, while a reticle pattern is superposed optically on the image by being reflected from the beam-splitter. In some types of two-beam interferometers, a beam-splitter separates an incident beam into two parts, each traversing a different path. Upon reflection by additional mirrors, these two beams reverse their paths and are combined to travel along a single path, each beam having undergone a reflection and a transmission at the beam-splitter. See INTERFEROMETRY.

Beam-splitters are conventionally produced by coating a surface of a transparent plate with one or more thin films of metallic or nonmetallic (dielectric) materials that transmit and reflect light of all wavelengths in the visible spectrum approximately equally so that both images have approximately the same color. The fraction of light transmitted or reflected depends on the thickness of the film, and can be accurately controlled in making the beam-splitter.

Metallic beam-splitting films absorb approximately one-third of the energy and transmit and reflect the remaining two-thirds; they are therefore not very efficient. Beam-splitters made from one or more layers of dielectrics are highly efficient, reflecting essentially all the energy not transmitted. These can be made to reflect all colors approximately equally, as a metal film, or to reflect certain desired colors more selectively than others. However the beam-splitter is made, the transmitted and reflected beams are complementary if no absorption occurs, their addition giving the incident beam. These beam-splitters may be used in color processes to separate an image into two or three component color images.

For some applications, the undesirable effects of a thick beam-splitting plate (multiple images, optical aberrations) may be overcome through the use of a pellicle. Pellicles are extremely thin transparent plastic sheets stretched over frames and coated with semireflecting films. Because of their extreme thinness (often less than 1μ), the multiple images are so closely superposed that they appear as a single image, and the aberrations introduced are negligible.

Spherical mirrors. Both concave and convex spherical mirrors may be used to form either real or virtual images. The image formed is not perfect, and the aberrations increase with the ratio of diameter of the aperture of the surface used to the radius of curvature of the sphere.

The position of the image with respect to the mirror can be determined from the position of the object and the radius of curvature of the mirror. The relationship is

$$\frac{1}{u}+\frac{1}{u'}=\frac{2}{R}=\frac{1}{F}$$

where u is the distance from object to mirror surface, u' is the distance from image to mirror surface, R is the radius of curvature of the mirror, and F is the focal length of the mirror.

Distances are measured along the line connecting the object and the center of curvature. According to one convention, for light from the object incident on the mirror from the left, object and image distances are positive to the left of the mirror, negative to the right. The radius of curvature of a concave mirror is positive, while that of a convex mirror is negative, because their centers of curvature lie respectively to the left and right of the mirror surface. Real objects and images lie to the left, virtual to the right. (A virtual object would form a real image to the right of the mirror if the mirror were removed.)

The ratio of the image height to the object height measured perpendicular to the line through the center of curvature is expressed as the lateral magnification M; $M = u'/u$.

If u and u' have like signs, the image is inverted; otherwise, it is erect.

Nonspherical mirrors. A spherical mirror produces an unaberrated image only at its center of curvature. Each of the other conic sections possesses a similar geometrical property that makes it a valuable optical element. For the paraboloid of revolution, an infinitely distant point source on its axis is perfectly imaged at the focus. This follows from the geometrical property that all rays emanating from the focus of a parabola are reflected parallel to the axis.

Ellipsoids of the prolate spheroid class, that is, those formed by rotation of an ellipse about its major axis, possess the property that a point source at one focus is, after reflection, perfectly imaged at the other focus. Hyperboloids of revolution produce a similar effect; the difference is that, for the ellipsoid, the object and image are either both real or both virtual, whereas for the hyperboloid they are opposites.

Astronomical telescopes usually employ one or more of the conics. The paraboloid may be used

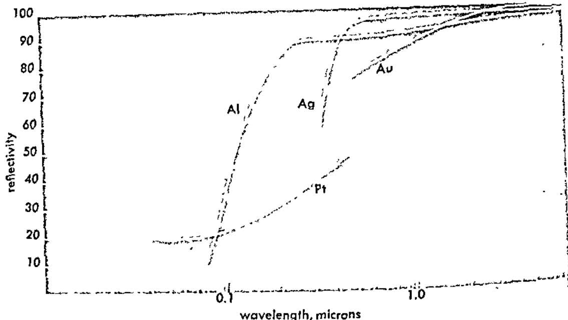


Fig. 3. Mirror reflectivity.

alone to form a Newtonian telescope. For systems of longer focal length, it may be combined with either a hyperboloid or an ellipsoid. In the Cassegrain telescope, a hyperboloid is placed in the converging beam formed by the paraboloid, one focus coinciding with that of the parabola, the other being located in the vicinity of the parabola. The paraboloid forms a virtual object for the hyperboloid which, after reflection, produces a real image.

In the Gregorian telescope, an ellipsoid placed beyond the focal point of the paraboloid has one focus coinciding with this point, the other near the paraboloid. Each system relays the primary image formed by the paraboloid to a secondary image at a much greater effective focal length.

Although these telescopes form unaberrated images on their axes (zero spherical aberration), they all suffer seriously from coma in the field. To improve this condition, the Schwarzschild-Cretien telescope employs two aspheric mirrors (with nonconic profiles) in the Cassegrain configuration to correct both spherical aberration and coma, increasing somewhat the field of good imagery. *See* TELESCOPE, ASTRONOMICAL; *see also* SPHERICAL AND ASPHERIC SURFACES, OPTICAL.

Mirror coatings. Almost all mirrors are produced by coating a correctly shaped glass surface with a thin film of highly reflective material. The process is usually carried out by vacuum deposition and the material selected according to the spectral region to be used.

For the visible region (0.35–0.75μ), silver provides the highest reflectivity, but aluminum is almost universally employed because of its much greater durability, brought about by the extremely thin, inert film of oxide that forms on exposure to air.

In the near infrared region, around 2μ, gold takes over, and maintains its superior reflecting properties throughout the infrared. However, the reflectivity of many metals in this region is almost equivalent to that of gold, so that unless many reflecting surfaces are required for a given instrument, the greater hardness of, for example, aluminum or nickel would recommend their use.

Aluminum provides the best reflectivity below 0.35μ, primarily because of a transmission region for silver around 0.31μ. Below about 0.10μ, aluminum becomes transparent and platinum is the highest reflecting metal (see Fig. 3).

The reflectivity values obtained for vacuum-deposited films depend markedly on the conditions of evaporation. In general, for high reflectivity and favorable aging properties, the substrate must be clean, the vacuum high (10^{-5} mm Hg), and the deposition rapidly produced. *See* REFLECTION (ELECTROMAGNETIC RADIATION). [H D PO]

Bibliography: F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3d ed., 1957; J. Strong, *Concepts of Classical Optics*, 1958.

Missile

An object capable of being projected or hurled, usually with the intent of striking some distant object. More particularly, a missile is usually a weapon that is self-propelled after leaving the launching device. The term thus excludes projectiles fired from guns, and free-falling bombs. The propulsion criterion is not an absolute test, however, since gliding bombs, especially those guided after launch, are also frequently classed as missiles.

Unguided and guided missiles. There are two primary categories of modern missiles; unguided and guided. All missiles, to be effective, must be

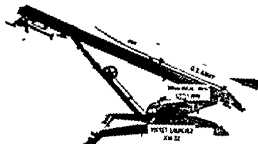


Fig. 1. Little John, an unguided, rocket-propelled missile. (U.S. Army)

directed in some sense, but are normally classed as unguided if they are subject to no further control after leaving the launching device (Fig. 1)

Typical of unguided missiles are the ground-launched and air-launched free-flying rockets used in enormous number in World War II.

The growth of the sciences of instrumentation electronics, and automatic control has led to the development of devices for the guidance of missiles in flight. Certain targets contrast with their surroundings by emitting or reflecting radiation in a distinctive manner. (Airplanes and ships are typical of the targets falling into this class.) The direction, and sometimes the distance, of targets of this kind can frequently be sensed by radiation receiving instruments. These instruments, located either in the missile or on the ground, can be used to guide the missile continuously toward the target. A guided missile with a high explosive warhead is frequently adequate to destroy a major target with the least expenditure of resources. The increase in accuracy of the guided missile over other means justifies the increased cost of the weapon.

Other targets cannot be readily distinguished from their surroundings by nonhuman means. Guided missiles used against these are frequently directed to the predetermined geographical location of such targets No terminal corrections of directional errors, such as are possible with the first class of targets, can be made As a result, guidance errors are likely to be substantial, particularly as the distance from the guidance station to the target becomes great In these cases, use of nuclear warheads becomes mandatory on economic grounds. Surface-to-surface missiles, particularly the ballistic types are typical of this class The decreasing cost and increasing versatility of nuclear warheads, coupled with their vastly superior destructive power, is causing them to displace conventional high explosives in all but the smaller missiles.

The main reason why long-range missiles can be more economical than manned bombers of similar performance is that the one-way nature of the missile permits a great reduction in size and cost. This size reduction would be unimportant if all bombers could return and be reused. In the face of the high attrition expected on long-range missions, however, the missile possesses a cost advantage.

Two other considerations have contributed to the growing importance of long-range missiles. These are the reduced expenditures of human life associated with pilotless weapons, and the increased ability of missiles to deliver highly concentrated blows

Propulsion requirements. Jet propulsion has contributed heavily to the ascendancy of the guided missile For missiles which spend all of their time within the denser atmosphere, air-breathing engines such as the turbojet and ramjet are frequently most advantageous (Fig. 2) Increased performance requirements are forcing the longer-range missile out



Fig 2. Regulus II, a guided, surface-to-surface aerodynamic missile which uses an airbreathing turbojet engine. (U.S. Navy)

into space where there is no air drag or friction heat. Out there, rocket propulsion is used exclusively.

Rocket propulsion, because of its simplicity and large thrust-to-weight ratio, is also frequently selected for shorter-range missiles, even though these travel entirely within the atmosphere (Fig. 3).

Guidance, jet propulsion, and nuclear warheads are becoming the standard ingredients for most modern missiles. With these elements, a long-range missile can deliver to a precise location, half a world away, a greater explosive load than the total of all explosives used in World War II. It can do this at an average speed of over 15,000 mph, reaching an altitude of nearly 1000 miles.

However, reliability limits a missile. A modern missile may be composed of 500,000 parts. Usually every one of these must function correctly if the missile is to perform its mission. Even after hundreds of missiles have been fired for developmental purposes, an over-all reliability of 100% may not be expected.

Guidance techniques. Although missile guidance systems are enormously complex, and perform many functions better than a human being, they do certain things poorly or not at all. Missile guidance systems are highly specialized. A missile that will seek out and destroy an aircraft is generally useless against surface targets. Target-seeking missiles have limited intelligence, they can be easily confused if the design of their seekers is known to the enemy. Multiple targets frequently cause a missile to have a "nervous breakdown" because it cannot decide which target to attack. Missiles cannot change their procedures on nearing a target if the situation is not as anticipated. They have no discretion.

Nontarget-seeking missiles require that precise target location information be supplied them prior to launching or during flight. Their accuracy is only as good as the information furnished. One cannot send such a device to look for targets of opportunity.

Missile classification. Missiles are classified in many ways. A common classification is according to the medium from which the missile is launched and to which it is directed. Thus, there are surface-to-air missiles, surface-to-surface missiles, air-to-surface, and so on. They may also be classified according to range (Figs. 4 and 5).

Missiles are also classified according to flight profile. The two categories are aerodynamic missiles (sometimes called cruise missiles) and ballistic missiles. Cruise missiles usually have wings or enlarged fins to give lift and maneuverability (Fig. 6). A ballistic missile has no wings (Fig. 7). It must be aimed sufficiently high to permit it to fall freely under the influence of gravity until it reaches the target (*see* BALLISTIC MISSILE). The elements of high speed, high altitude, and free fall or ballistic trajectory characterize most recent long-range missiles including intercontinental and intermedi-

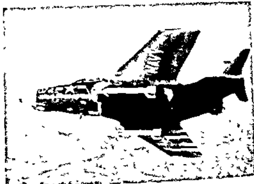


Fig. 3. Sidewinder, an air-to-air, solid-rocket-propelled guided missile. (*U.S. Navy*)

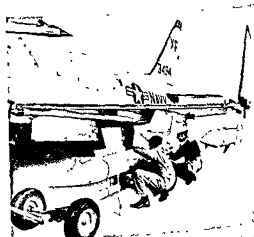


Fig. 4. Bullpup, a short-range, air-to-ground, solid rocket-propelled guided missile. (U.S. Navy)

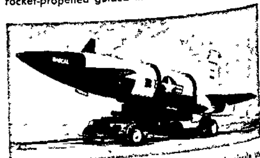


Fig. 5. Rascal, a long-range, air-to-ground missile using a liquid propellant rocket engine. (U.S. Air Force)



Fig. 6. Snark, an intercontinental cruise type missile which uses an air-breathing turbojet engine. (U.S. Air Force)

ate-range ballistic missiles. Long-range ballistic missiles may not even have fins, but may depend on jet reaction for stability and directional control. It is common practice to swivel the rocket engines of large ballistic missiles. This action provides adequate steering moments for control even though the missile may be aerodynamically unstable (*see* GUIDED MISSILE).

Missile subsystems. Guided missile systems are also frequently discussed according to the subsystems which go to make up the whole. The most common breakdown is as follows: airframe, guidance, propulsion, payload, auxiliary power, and ground support.

Airframe. A missile airframe usually includes the primary body structure, wings, and fins. Propellant tanks or casings are sometimes considered part of the airframe, and sometimes part of the propulsion system. Functionally, they may be part of both subsystems, particularly if they are integral with the skin of the missile body.

Guidance and control. The term guidance sometimes is used to include both guidance and control. When the two terms are used separately, guidance is limited to equipment that determines the difference between the actual flight path and the desired path, and control pertains to equipment that takes the error signal from the guidance equipment and steers the missile back to the proper path. Thus, the guidance subsystem might consist of radar and computers, either missile-borne or ground-based, which track the missile (and sometimes the target) and determine the proper direction and distance for the missile to go. The control subsystem would consist of an autopilot and an amplifying servoactuator. The autopilot would contain gyroscopes to maintain a steady reference when no guidance signals are received. The servoactuator takes a weak electrical signal from the guidance subsystem and converts it into a powerful mechanical force for operating control surfaces or jets.

Propulsion. The propulsion subsystem usually consists of main and auxiliary propulsion engines, the propellant or fuel feed equipment, the devices for maintaining pressure in the propellant tanks (in liquid-fuel rockets), and sometimes, the propellant tanks themselves. Turbojet engines, ramjets, liquid- and solid-fuel rockets have all been used in modern missiles. Occasionally, two different types of engines are used on the same missile (Fig. 8). Quite generally, missiles with air-breathing engines for sustained propulsion use rockets for quick initial acceleration; these rockets are called boosters. The term has been extended to include the first stage of multistage rockets, even when this stage comprises over 75% of the total mass of the missile.

Choice of propulsion subsystem is dictated primarily by two factors: speed and range. High speed within the atmosphere requires large thrust. Engines which give high thrust for their size and weight usually have high specific propellant con-

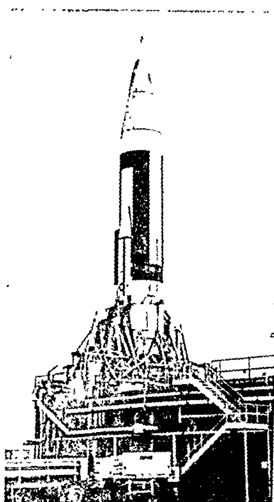


Fig. 7 Atlas, intercontinental ballistic missile similar to Titan. Atlas also uses liquid rocket engines. (U.S. Air Force)

sumption, or, to use the term more commonly used in rocketry, low specific impulse.

$$\text{Specific impulse} = \frac{\text{thrust} \times \text{time}}{\text{propellant burned}}$$

If the speed is high and the missile range is short, the burning time is also short; hence, the high consumption of propellant per unit of time will not overshadow the saving in engine weight. In contrast, if range is long and speed is low, an engine having large weight per unit thrust but lower fuel consumption will prove superior.

The never-ending contest between offense and defense forces weapons for both purposes to higher and higher speeds. As the speeds become very high, the high air drag and extreme heat of friction force missiles to travel at progressively higher altitudes. An altitude is finally reached, depending on the state-of-the-art, where air-breathing engines do not produce enough thrust to overcome the drag. Rocket engines operate at any altitude. If the rocket can

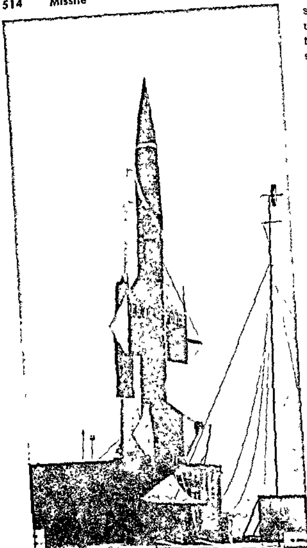


Fig. 8. Bomarc, a surface-to-air guided missile. This interceptor missile uses two types of engines, liquid rocket for boost and ramjet for sustained thrust. (*U.S. Air Force*)

travel at a sufficiently higher altitude than the air-breathing engine, the reduction in drag more than compensates for the higher specific propellant consumption of the rocket. Under these conditions, the aerodynamic forces are insufficient to sustain the missile in level flight and a ballistic path must be followed.

Payload. The missile may serve to carry instrumentation along its flight path (*see* METEOROLOGICAL ROCKETS) or to deliver a warhead to the distant target. The payload housing protects the payload during flight.

High-speed missiles present problems of protecting sensitive components from the high temperatures produced by aerodynamic drag and sometimes from high accelerations. The speed of long-range ballistic missiles, in particular, is so high that special heat protection is required. The problem is much alleviated by the fact that the trajectories of such ballistic missiles are almost entirely in a vacuum. Only during the reentry phase is heat protection necessary (*see* REENTRY). This phase is of short duration, normally less than a minute.

The shape of the reentering body is made quite blunt (Fig. 9). This shape causes a greater fraction of the kinetic energy of the missile to be transferred to the surrounding atmosphere. The remaining kinetic energy reappears as heat generated on the missile surface. One of two means is provided to keep this heat from damaging the payload. The first is simply to supply a sufficient mass of material to absorb the heat without an excessive rise in temperature. This is the heat-sink method. The second is to use a protective shield made of a material which melts or sublimes. This action absorbs heat. The melted or gasified material is removed by the air friction. This technique is called ablative protection; it appears to permit lighter payload housings, but it is more difficult to treat theoretically and thus may require more cut-and-try development. *See* NOSE CONE.

Nuclear warheads are being used increasingly. Improvements in nuclear technology have brought about a continuing decrease in the size and cost of atomic explosives. The increase in effectiveness that atomic warheads give even to antiaircraft missiles is sufficient to justify their additional cost compared to chemical warheads. Smaller nuclear warheads use fission of plutonium or uranium 235 as the source of energy. Explosive yields ranging from

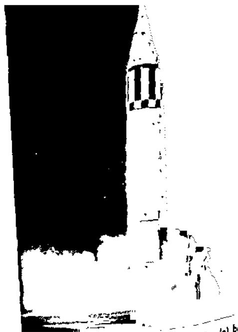


Fig. 9. Jupiter, intermediate-range (1500-mile) ba tic missile propelled by a liquid rocket engine equipped with a reentry type of nose cone (*Army*)

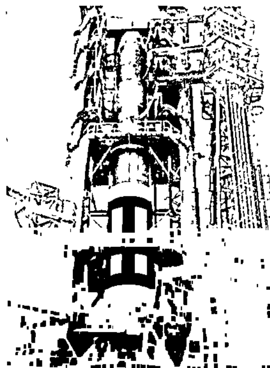


Fig. 10. Titan in its servicing tower, being prepared for flight. (*U.S. Air Force*)

1000 to over 100,000 tons of TNT are attainable with fission. For higher yields, combinations of fission and fusion reactions are used. A fission reaction produces the high temperatures necessary to produce fusion of lighter elements such as hydrogen or tritium. The major portion of the energy obtained with the larger warheads comes from the fusion process. Fusion can produce an energy release limited only by the size of the bomb. The equivalent of millions of tons of TNT has been obtained.

Many different types of fuzing are used. The warheads of antiaircraft missiles are frequently detonated by a proximity fuze, which senses the closest point of approach or the presence of the target within a lethal range and direction. Warheads used against surface targets usually are detonated by an altitude-measuring device. Sometimes an impact fuze is provided as a backup in case the altitude fuze fails.

Ground support. Included in ground support equipment (GSE) are all elements of the ground-based portion of the missile system except, strangely enough, the ground portions of the guidance system. Other elements include transporters, missile checkout equipment, and propellant servicing equipment (Fig. 10). Adjacent to the launching site are consoles of instrumentation and control to fire the missile and monitor its behavior prior to launching and, for experimental work, during flight.

Extremely elaborate devices have been developed for determining whether an assembled missile is in proper working order. Some of this equipment not only checks each system automatically in sequence, but isolates failures and even checks for malfunctions within the checkout equipment itself. *See* LAUNCHING PAD COMPLEX.

It is so much easier to develop ground-based equipment than compact lightweight equipment carried within the missile that there is a strong tendency to expand the ground portion of the system if, by so doing, the missile can be simplified even slightly. This trend is in contrast to the usual practice with aircraft, where the ground support equipment plays little part in preflight checkout and in-flight monitoring of vehicular performance. [R.C.T.]

Missile guidance systems

Equipment for directing a missile to its target. In general, missile guidance systems must (1) locate and track the target, (2) predict future target motion, (3) compute an interception course, and (4) provide the missile's autopilot with steering signals. For discussion of the autopilot portion of the systems *see* AUTOPILOT.

Guidance methods depend upon the mission and characteristics of the missile. Short-range missiles may be guided either by a command system from the ground, or by a homing system in which the missile carries its own target-seeking radar.

For longer-range missiles, guidance is a navigation problem, and radio, radar, or inertial navigation is used.

Ballistic-missile guidance is confined primarily to the boost stage, and may be performed either by an inertial system or a ground command system.

Short-range missiles. A ground command system (Fig. 1) is advantageous for short-range mis-

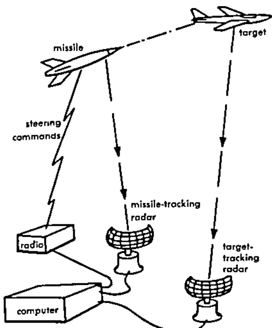


Fig. 1. Command guidance system.

siles because the missiles carry only a radar receiver and an autopilot; the computer is on the ground and may be as sophisticated as desired. The range of the system is limited, as is also the accuracy with which it can resolve geometric angles involved in the missile-target tactics.

These disadvantages are less severe for air-launched missiles where the command system is carried on board the nearby launching aircraft.

In contrast, radar or infrared equipment in the homing missile itself measures the relative motions between target and missile directly. Measurements are more accurate from the missile's vantage point, and computations based on relative motion are simpler. Because the missile must carry the additional load of a seeker and computer, the success of the system depends on the degree to which these subsystems can be miniaturized and simplified.

In an active homing system the missile's own radar illuminates the target, thus making the missile entirely autonomous, once the target has been "captured" by the missile's radar.

In a passive homing system the missile carries only a radar or infrared receiver, and the target must be illuminated from a separate transmitter. (An infrared system is inherently passive because it depends only upon heat radiation, usually from the engine of the target.)

A beam-rider system is even simpler than the passive seeker, a radar receiver in the missile simply seeks to keep the missile in the beam of a ground radar. The ground station in turn keeps the radar beam centered on the target, so that the missile is guided along the line-of-sight to the target.

The accuracy of a guidance system for short-range missiles depends upon the accuracy of the sensor employed, the sophistication of the computer, the tactical capability of the missile, and the reliability of every component.

Radar glint noise. The performance of systems using a radar seeker is limited by radar glint noise. The phenomenon occurs because the radar echo

fully about the sky through distances comparable with its over-all dimension. These fluctuations in position must be smoothed (filtered) so that the missile tracks an average point. Filtering involves time delay so that, if filtering is too heavy, the missile may follow the target maneuvers too slowly, and if filtering is too light, the missile may hit a ghost target or may expend its maneuvering capacity unnecessarily. Glint noise is inherent and random in nature; hence system accuracy must be studied on a statistical basis.

Missile tactics. The geometry of air-to-air interception as indicated in Fig. 2 provides several possible techniques for steering the missile into collision with the target.

In the pursuit course the missile is steered so that its velocity vector is always along the line-of-

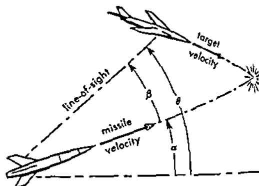


Fig. 2. Interception geometry.

sight to the target. Mathematically, it is desired to keep angle β equal to zero; thus, the steering command is $\alpha_c = C\beta$ in which C is a control constant and α_c is the input to the autopilot rate control. That is, the missile is turned at a rate proportional to its deviation from the required condition that $\beta = 0$. In this respect, a beam-rider system inherently involves a line-of-sight or pursuit attack.

The pursuit mode of control is simple to mechanize and is most successful in head-on or tail chase attacks. However, if the attack is from the side of the target, the line of sight may be turning so rapidly that the missile cannot follow.

In a collision or constant-bearing course, the missile is maneuvered to keep the line of sight at a constant angle in space. Mathematically, θ remains fixed. The steering command in this case is of the form $\alpha_c = C\dot{\theta}$.

Proportional navigation, a more general technique, may involve a combination of both the above steering equations.

Second-order prediction may be employed to improve the interception course still further. The computer predicts both the target's future position (as, in effect, it does in any method), and its future velocity. The effect of glint noise is especially severe in this scheme.

airplane, a double

g the airplane

g the missile to

the target.

The selection of the mode of attack, together with values for guidance parameters, is based on statistical studies of the probability of hitting the target, termed kill probability. Kill probability is maximized by the most efficient tactical use of missile velocity, maneuvering capacity, and optimum filtering of glint noise. Hours of flight simulation may be necessary to obtain sufficient statistical data for selection of guidance parameters.

Long-range missiles. Guidance of long range (nonballistic) missiles is a problem of geographic navigation. In special cases a series of ground stations may form a command guidance system. Usually, however, the navigation system is carried in the missile, and guides it along a programmed course to a geographic destination.

Radio equipment can be used to navigate a missile automatically through a network of ground stations arranged in a long-range radio-navigation system (see LORAN). Radar may be used for continuous map-matching or to establish check points. Doppler radar velocity measurements can be used as the basis for continuous dead reckoning. Both radio and radar systems have the disadvantage that they emit radiation, which can be detected and jammed or distorted by enemy countermeasures. See ELECTRONIC COUNTERMEASURES.

The most versatile long-range guidance system is the inertial autonavigator, which is completely self-contained and emits no radiation. A pure inertial system simply measures absolute vehicle motions and continuously computes distance and direction moved from the launch point. The system may be augmented by an automatic star tracker to obtain celestial fixes, by periodic check-point information, or by a Doppler radar to give auxiliary velocity information. See INERTIAL GUIDANCE SYSTEM.

Ballistic missiles. A ballistic missile is guided precisely along a given course during the time it is being propelled by its rocket engines (at most, a few minutes). After the engines are cut off, the missile follows a free-flight trajectory to the target, which may be several thousand miles from the launch site. The boost path may be likened to a gun barrel so that the flight direction (aim) and rocket cut-off velocity (muzzle velocity) are all-important to the accuracy of the shot. See BALLISTIC MISSILE.

Guidance of the ballistic missile during its powered flight can be accomplished either by precise command guidance using a series of tracking stations or, alternatively, by means of a special-purpose, short-time inertial navigation system capable of accurate performance under high acceleration. See GUIDED MISSILE. [R.H.C.]

Mississippian

A large division of late Paleozoic geologic time, varyingly considered to rank as an independent period or as a main subdivision (termed epoch) of the Carboniferous Period. It is named Mississippian for a succession of highly fossiliferous marine strata consisting largely of limestones found along the Mississippi River between southeastern Iowa and southern Illinois. The rocks formed during the time defined as Mississippian are classed as a geologic system by most American geologists and this practice is increasingly accepted outside of North America even though foreign usage favors Lower Carboniferous. Widespread marine transgression of the continents, accompanied especially by accumulation of limestones, characterized Mississippian time. Locally, this was preceded or succeeded by mountain building. See CARBONIFEROUS. [R.C.M.]

Mistletoe

This is the name given to several species of the mistletoe family (Loranthaceae), especially those of the genus *Loranthus*. The true mistletoe of Europe is *Viscum album*, and among the early nations this was an important ceremonial plant,

Mistletoe (Phoradendron flavescens). (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

which probably accounts for the origin of the custom of kissing under the mistletoe. In the United States, the common representative of the group is *Phoradendron flavescens*. All of the mistletoes are green hemiparasites; that is, they obtain water and minerals from the host plant but manufacture their own food. See SANTALALES. [P.D.S.]

Mite

Any of numerous small species of the order Acarina, class Arachnida, phylum Arthropoda. Mites and ticks are members of the same order; they differ primarily in size, the mites being the smaller members of the order. Most mites are quite small, many of them microscopic in size.

Mites and ticks differ from other Arachnida and in fact from all other arthropods in having a saclike body, with the abdomen not distinctly segmented, and broadly joined to the cephalothorax.

Whereas mites live in a variety of habitats and have very diverse feeding habits, a substantial number are predators or parasites and attack many different kinds of plants and animals. Several species are of economic importance.

One of the best known mites is the chigger or red bug. The itch mite is another well-known parasite of man. Other varieties of the same species, *Sarcoptes scabiei*, attack dogs, hogs, horses, and other mammals, causing the disorder commonly called mange.

Human itch is one of the oldest known maladies of man. The female itch mite, about 1⁄60 in. in di-

Mite. (a) Itch mite, *Sarcoptes scabiei;* length 1/50 in. (b) Chicken mite, *Dermanyssus gallinae;* length to 1/35 in. (*From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

ameter, deposits 10–25 eggs in a burrow it has made in the skin. The eggs hatch quickly and the young attain the adult stage after three molts, accomplished in about 10–12 days. The males are about half as large as the females.

The two-spotted spider mite, *Tetranychus telarius,* is perhaps the most widespread and destructive of the mites attacking plants. It sucks juices from a wide range of plants, causing the leaves to turn brown and fall.

Young chickens and

ens of all ages, adversely affecting their growth and general health. Chicken mites bite through the skin and feed on the blood of the host. *See* ACARINA; CHIGGER. [J.D.B.]

Miticide

A chemical useful for killing phytophagous and parasitic members of the order Acarina of the class Arachnida. Although these animals are not insects, their control has become a part of economic entomology. A major difference between the mites and ticks and insects is that mites and ticks have eight legs, whereas insects have only six. The most apparent difference between mites and ticks is that the former are usually very small in size, whereas the latter are quite large and range up to ½ in. or more in length. All economically important species of ticks are parasitic. Several species of mites are parasitic, but a considerable number are economically important as plant pests. Chlorinated hydrocarbon insecticides, particularly DDT, benzene hexachloride (γ-isomer), and toxaphene, have proved useful for the control of parasitic ticks and mites, but have little or no value in the control of phytophagous mites.

Miticides are chemicals used specifically for the control of plant-feeding mites. Among the more effective chemicals available are: di-(*p*-chlorophenyl)methylcarbinol, Kelthane [2,2-bis-(*p*-chlorophenyl)-2-hydroxy-1,1,1-trichloroethane], Chlo-

tetrachlorodiphenyl sulfone), Chloroparacide (*p*-chlorophenyl *p*-chlorobenzyl sulfide), Sulfenone (*p*-chlorophenyl phenyl sulfone), and Aramite [2-(*p*-*tert*-butylphenoxy)-isopropyl 2-chloroethyl sulfite]. All are specific miticides and have little or no insecticidal activity. Many organic phosphorus insecticides, however, are also quite effective as miticides. *See* ACARINA; INSECTICIDE; PESTICIDE. [C.A.H.; C.F.L.]

Mitochondria

Cell structures, sometimes called chondriosomes, which are minute cytoplasmic organelles found in all cells. They are clearly visible in the living cell when examined with the dark field or phase contrast microscope, and can be identified with the bright field microscope when specifically stained by Janus green, applied supravitally. They may appear as small spherical granules, short rods, or long filaments (Fig. 1). The spherical forms range from 0.2 μ to more than 1.0 μ in diameter and filamentous forms may vary from 1 to 7 μ in length. Although the shape of the mitochondria tends to be uniform and characteristic of the particular cell type, they are quite labile and responsive to alterations in their intracellular environment. Thus they may change form in different physiological conditions and are sensitive indicators of cell injury. In living cells observed in vitro they are in constant motion, executing slow flexuous movements as they move from one part of the cell to another. As a rule they maintain their characteristic shape dur

Fig. 1. Drawing of intestinal epithelial cells showing numerous filamentous mitochondria concentrated in the apical cytoplasm.

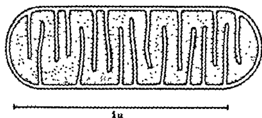


Fig. 2. A diagrammatic longitudinal section of a mitochondrion as revealed by electron microscopy.

ing these movements, but occasionally filamentous forms are observed to break up into short rods or spheres. In other instances they elongate and thrust out lateral branches which coalesce temporarily with other parts of the same mitochondrion to form annular or plexiform (network) structures. Mitochondria of such bizarre configuration are relatively uncommon, but their occurrence emphasizes the variability and plasticity of this cell organelle.

The origin of new mitochondria is obscure. Whether they arise only by binary fission or segmentation of preexisting mitochondria, or whether they can arise initially from the cytoplasmic matrix is still a matter of dispute.

Submicroscopic structure. In electron micrographs of thin sections of cells the mitochondria are seen to be limited by two dense lines ~50 angstroms (A) thick and ~50 A apart (Fig. 2). The outermost of these is smooth and continuous, whereas the inner turns inward at more or less regular intervals along the length of the mitochondrion to form a series of thin lamellae, the cristae mitochondriales, that project into the interior of the organelle. According to the currently accepted interpretation of the electron microscope images, the mitochondrion is limited by two separate membranes, the innermost of which is infolded to form the cristae. An alternate interpretation that has not gained general acceptance holds that the parallel dense lines that bound the mitochondria in electron micrographs and the clear space between them represent components of a single membrane with a trilaminar molecular organization composed of a layer of lipid flanked by two layers of protein. The proponents of this interpretation also visualize the cristae as a series of single membranes which form nearly complete transverse septa not necessarily continuous with the limiting membrane. *See* LIPID; PROTEIN.

Although the internal membranes usually take the form of folds or cristae, exceptions are found in the mitochondria of protozoa and in certain cell types of higher animals where they form slender villous or tubular projections called the tubuli mitochondriales. *See* PROTOZOA.

The interior of the mitochondrion is occupied by an amorphous matrix of appreciable density. A few dense granules ~500 A in diameter may be scattered in the matrix, but these are not found in all cell types and apparently are not essential.

The basic pattern of structural organization is the same for mitochondria of animal and plant cells in general, but there are marked variations from one cell type to another in the number of cristae or tubuli per mitochondrion. Large numbers are characteristic of tissues of high metabolic activity, whereas fewer are found in the mitochondria of cells having lower energy requirements (Fig. 3). Much of the enzymatic activity of the mitochondria resides in their membranes and the cristae and tubuli appear to be the means of increasing the area of active surface exposed to substrates in the matrix.

Number and distribution. No accurate information is available on the numbers of mitochondria per cell. In early embryos most of the cells contain approximately the same number, but as development proceeds and the tissues become specialized, distinctive differences in number and shape become apparent. Their distribution usually appears to be random, but in some cell types they establish special topographical relations to other organelles and inclusions. For example, they often gather around the centrosome (*see* CENTROSOME). In intestinal epithelium, they tend to congregate in the apical cytoplasm, whereas in the renal epithelium they are most abundant at the base. In muscle mitochondria are intimately associated with the Z bands of the myofibrils and in the spermatozoan they wrap around the base of the flagellum. *See* MUSCLE.

Composition and function. In 1934 mitochondria were isolated in relatively pure form by differ-

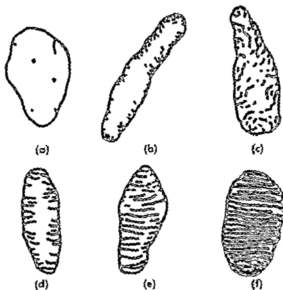


Fig. 3. Drawings of representative mitochondria from electron micrographs of several tissues illustrating variations in the form and extent of internal membranes. (a) Bronchiol epithelium. (b) Interstitial cell of testis. (c) Paramecium. (d) Liver. (e) Pancreas. (f) Cardiac muscle.

ential centrifugation of saline suspensions of tissue homogenates. The method was subsequently modified by the introduction of sucrose solution as the suspending medium during centrifugation. This improvement made it possible to maintain nearly normal morphology of the mitochondria during isolation and thus to retain much of their original chemical activity. Since this development, mitochondria have been the subject of intensive biochemical investigation. About 65% of their weight is protein and 35% is lipid of which almost two-thirds are phosphatides (*see* PHOSPHATIDE). They contain highly integrated systems of enzymes providing energy for cell metabolism. Among the most important of these systems is the tricarboxylic acid cycle and its associated respiratory enzymes which produce the energy-rich compound adenosine triphosphate which is essential for a number of vital cellular activities, including muscular contraction. *See* ADENOSINETRIPHOSPHATE (ATP). In the course of their assimilation, all three classes of nutrients —protein, carbohydrate, and lipid—can enter the carboxylic acid cycle through acetyl coenzyme A or some other intermediate and can be utilized either for energy production or as substrates for the synthesis of new protoplasm or cell products (*see* PROTOPLASM). Enzymes that participate in synthesis of phospholipids and in synthesis and degradation of fatty acids are located in the mitochondria as well as some components of the uric acid cycle. Correlated biochemical and electron microscopic studies suggest that the mitochondrial enzymes are mainly associated with the membranes and that their arrangement may be important for maintaining the appropriate spatial relation of the enzymes and substrates involved in the highly integrated reaction sequences characteristic of the living cell. *See* ENZYME. [D.W.F.]

Bibliography: *See* CELL (BIOLOGICAL).

Mitosis

The term, as first used by W. Flemming in 1882, applied to a division of the cell nucleus in which a spindle and chromosomes are involved. Mitosis is still used in that sense by cytologists, but most embryologists use it loosely to cover cell division as a whole. On the other hand, geneticists have come to restrict the term to the division process as it occurs in somatic cells and germ cells before maturation, using *mitosis* as an antonym to meiosis, the division process seen in maturing germ cells. This is an unfortunate use of the word mitosis because the division process in meiotic cells involves mitosis to the same extent that it does in somatic cells. *See* MEIOSIS.

Spindle. The primary apparatus in mitosis is the spindle, a bicuspid structure, which in many forms, especially among animals, carries an astral configuration at each of its two poles. It is through their interaction with this apparatus that the chromosomes, the essential constituents of the nucleus, are transported to form two daughter nuclei. The spindle with its asters, when fully formed at the time of metaphase, constitutes a discrete and distinct structure. It can be seen without difficulty in the normal living cell and the fact that it can be moved as an intact body with a microdissection needle or by centrifuging long ago convinced cytologists of its physical distinctiveness from the other cell constituents. The spindle has, after certain pretreatments, been separated from the rest of the cell contents.

Doubts have long existed concerning the reality of the various structural elements that can be seen within the spindle after good fixation (Fig 1). Most of these elements cannot be seen with certainty in the living spindle and for many years cell

work of S. Inoué using a modified polarization microscope has demonstrated their reality in the living cell and has thus terminated a long series of arguments. *See* MICROTECHNIQUE.

Judging from existing cytological evidence, the spindle is in most cases formed from materials in the nucleus. In many cases the spindle makes its appearance long before the nuclear membrane breaks down, and in a few forms, for instance, in certain Protozoa, this membrane remains intact throughout the entire mitotic cycle. This does not mean that the cytoplasm may not sometimes contribute to the spindle structure; and it is of course conceivable that certain constituents of the cytoplasm may enter the nucleus even when its membrane is never broken down.

The spindle and its composing elements cannot be considered as an apparatus that is fully constructed before the chromosomes are in some way attached to it for their transport to the poles. Instead, it is quite certain that the chromosomes,

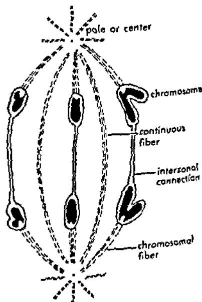


Fig. 1. Structural elements of the mitotic spindle

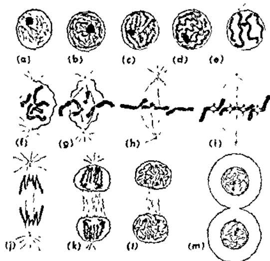


Fig. 2. Diagram of the different stages of mitosis. (a) Interphase. (b), (c), (d), (e) Prophase, in which there is progressive contraction and condensation of the chromosomes (each one composed of two chromatids). (f), (g) Prometaphase; the spindle is beginning to form and the nuclear membrane is disappearing. (h), (i) Metaphase. (j) Anaphase (k), (l), (m) Telophase. The centromere is indicated by a clear circle in each chromosome. (E. D. P. De Robertis, W. W. Nowinski, and F. A. Saez, General Cytology, 2d ed., Saunders, 1954)

and more especially their kinetochores (or centromeres), participate in the formation of some of the fibrous tracts in the spindle and thus play a role which may be as important as that played by polar forces in the construction of the spindle apparatus.

Chromosomes. The chromosomes, enormously extended during interphase, contract by spiralization during prophase and are most sharply defined at metaphase and anaphase. It is at metaphase that the chromosomes are assembled in a flat configuration or equatorial plate while the nuclear membrane breaks down and the spindle makes its appearance. During the anaphase which follows, the two halves of each chromosome are transported to opposite poles where, during the telophase, they are surrounded by a new nuclear membrane and again assume the interphase condition (Fig. 2).

The halving of each chromosome is a process quite distinct f...

is a reduplication of the essential constituents of the chromosome after division, and the primary activity of the spindle is thus only the transport of such halves or daughter chromosomes into the two new cells that are formed from the parent cell.

This superficially simple process of transportation is in reality exceedingly complicated and after many years a fully satisfactory explanation for it is still not available. Various hypotheses have ranged from a pulling or traction mechanism, to viscosity changes, diffusion, hydrostatic forces, and many others, including that of considering the spindle as a liquid crystal or tactoid with all the forces that may be available there. The hypothesis perhaps most favored at the present time represents a return to the old idea that the chromosomes are pulled to the poles by means of the chromosomal spindle fibers. It is considered likely that all the spindle fibers originate through an orientation of long protein molecules into longitudinal tracts and that the folding of protein chains can exert sufficient traction on an attached chromosome to take it to the pole.

This concept presents minor difficulties, of which the mode of attachment of such fibers to the chromosome on the one hand and the pole on the other is not the least. Again, this most obvious part of the mitotic cycle, the transport of the chromosomes to the poles, involves several preparatory steps that are possibly even more difficult to analyze. Thus there is still the puzzling question of how all the chromosomes come to assume an equatorial position before the movement to the two poles begins.

Many such questions will no doubt receive natural answers when cytological information is bolstered by an increase in knowledge of the physical chemistry of the cell. It is very likely that more than one of the older hypotheses will be found to be partially correct, even though some modification of its original formulation may be necessary. See CELL DIVISION; CHROMOSOME; CYTOKINESIS; CYTOLOGY. [F.SC.]

Bibliography: A. Hughes, *The Mitotic Cycle*, 1952; F. Schrader, *Mitosis, the Movements of Chromosomes in Cell Division*, 2d ed., 1953.

Mixer

A device having two or more signal inputs, usually adjustable, and one common output. It is used in audio amplifiers to combine the outputs of individual microphones and other audio-signal sources linearly and in desired proportions to produce one audio output signal (see MICROPHONE). In television, a mixer serves similarly to combine the outputs of two or more television cameras or other video signal sources. The mixer stage in a superheterodyne receiver combines the incoming modulated rf signal with the signal of a local rf oscillator to produce a modulated i-f signal (see RADIO RECEIVER). Crystal diodes are widely used as mixers in radar and other microwave equipment. [J.MR.]

Mixing

An operation to bring about distribution, intermingling, and homogeneity of matter. Differences in the type and amount of mixing can affect such

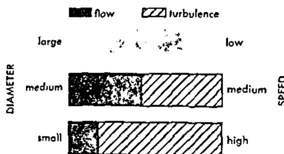


Fig. 1. Effect of impeller size and speed on flow and turbulence at constant power.

various processes can be regulated. Mixing is accomplished by a great variety of equipment.

Solids are mixed with solids in rotating cylinders or similar equipment. Most fluid mixing is done by rotating impellers that operate in cylindrical vessels. The impeller is driven by a shaft attached to a power transmission device connected to an electric or other motor. The mixer produces mechanical effects only Molecules of themselves will diffuse, but mixing impellers produce flow which results in forced convection. Hence, reactants can be brought to an interface as rapidly as desired by controlling the forced convection currents with mixers.

Fluid mixing. There are five distinct types of operations in which fluid mixers are used: blending of miscible fluids; stirring of immiscible liquids for extraction, emulsification, and other processes; suspension and dissolving of solids; agitation of gases and liquids; and heat transfer These may be batch or continuous operations. Mixer performance criteria will be different for each, but the fluid motion produced by the impeller and system is the controlling factor.

Fluid motion, both large-scale (mass flow) and small-scale (turbulent), are ordinarily required to bring about rapid mixing. The discharge *stream* from an impeller initiates the large-scale flow pattern. Turbulence is generated mostly by the velocity discontinuities adjacent to the stream of fluid flowing from the impeller, and also by boundary

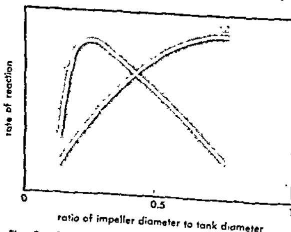


Fig. 2 Effect of impeller size on reaction rate at equal power input.

and form-separation effects. Turbulence spreads throughout the mass flow and is carried to all parts of the container. Some mixing operations require relatively large mass flows for best results, whereas others require relatively large amounts of turbulence. There is usually an optimum ratio of flow to turbulence for a desired mixing operation, whether it is a simple blending of immiscible liquids or a mass-transfer operation followed by chemical reaction.

Energy must be supplied to produce fluid motion. Thus, in comparisons of mixing with different equipment or with different sizes of the same type impeller, it is essential that equal power be supplied at the impeller. For the same power, the ratio of flow to turbulence from mixing impellers can be changed. Figure 1 illustrates the differences in mass flow and turbulence which can be achieved for the same power input for any one shape of dimensionally similar impellers. A large-diameter, low speed impeller produces a large ratio of flow to turbulence, whereas a small-diameter, high-speed impeller will give a small ratio.

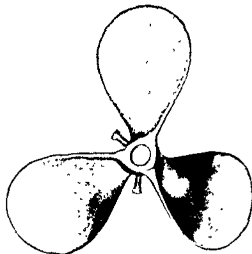


Fig. 3. Propeller. (*Mixing Equipment* Co.)

Curve A, Fig. 2, illustrates a reaction best accomplished by large flow and small turbulence. The ratio of impeller diameter to tank diameter in this figure is proportional to the flow to turbulence ratio. Curve A, typical of blending operations, shows that the rate of blending increases to a maximum as impeller diameter is increased (and impeller speed is decreased) with power input constant. Curve B of Fig. 2 is typical of gas-liquid mixing operations. Here the rate of mass transfer between phases increases to a maximum at small impeller diameter and then decreases as impeller diameter is increased. The significance is that more turbulence is available with the small *impeller* and that turbulence is more important than flow in this operation.

In all bench-scale and pilot-plant work where mixing is important, the effect of the impeller diameter to tank diameter ratio should be determined so

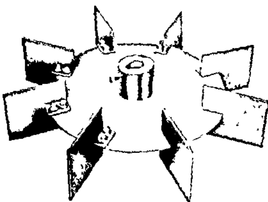


Fig. 4. Flat-blade turbine. (*Mixing Equipment* Co.)

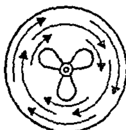


Fig. 5 Swirling flow pattern for impeller of any shape, without baffles.

that the type of flow motion best suited to the operation can be found. If an optimum ratio is found, it becomes the basis for larger-scale design.

Mixing vessels and impellers. Flow motion is dependent upon the shape and fittings of the container, the shape and position of the rotating impeller, and the physical properties of the fluid The best mixing usually produces lateral and vertical flow currents, and these currents must penetrate to all portions of the fluid.

Cylindrical vessels are the best vessels for mixing. For the laboratory, it is recommended that beakers or resin-type flasks be used whenever mixing is of importance. Round-bottom flasks are the [illegible] undrical glass or metal containers should be provided with vertical baffles, or if a propeller is used in the off-center position the baffles may be omitted.

The most useful impellers are the marine-type propeller (Fig. 3), the turbine (Fig. 4), and the simple flat paddle. The use of any of these on a vertical shaft rotating on the center line of a cylindrical vessel, produces rotating fluid motion (Fig. 5). A vortex forms, around which the liquid swirls. This motion often results in separation or stratification rather than mixing. A minimum of turbulence and of vertical and lateral flow motion will result. Very little power can be applied. Rotary motion (and surface vortex) can always be stopped by inserting projections into the body of the fluid, and the flow motion will be decidedly different. When these projections are at the side of the tank they are called baffles, and are commonly used to prevent rotary motion and to obtain good mixing in large-size industrial equipment. The propeller with tank wall baffles (Fig 6*a*) will produce an axial flow pattern, and the paddle and turbine will produce radial flow as shown in Fig. 6*b*. These flow patterns cause a maximum of lateral and vertical flow motion and are conducive to good mixing.

Probably the most convenient technique in the laboratory to avoid swirl is to use the propeller off-center. A propeller discharging downwards and positioned precisely as in Fig. 7 (note the direction of rotation and position) will result in good vertical and lateral flow motion without swirl and surface vortex; the position is critical but is easily found by trial

Liquid in large tanks (3,000,000 gal., for example) is often blended by use of a side-entering propeller mixer. Proper angular positioning is imperative; Fig. 8 shows the correct position for a right-hand, clockwise-rotation propeller.

Low-viscosity liquids behave as shown in the illustrations. High viscosity liquids (above 1000 centipoises) will have much less rotary flow, and the motion will approximate that in Fig. 6 even if baffles are not present.

Mixing in small-scale laboratory work should be done with one of the three types of impellers mentioned, they should be scale models of the large industrial ones. Of the thousands of different impeller shapes available, it is difficult to find any that are better than these three forms. Most large-scale in-

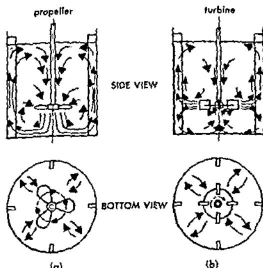


Fig. 6. (a) Flow pattern for propeller with baffles at tank wall. (b) Flow pattern for turbine with baffles at tank wall.

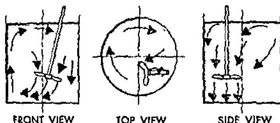


Fig. 7. Flow pattern for top-entering, off-center propeller without baffles.

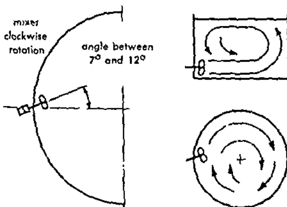


Fig. 8. Side-entering propeller mixer position.

stallations use either propellers or turbines, and the size and speed can be related directly to the laboratory models for equivalent performance.

Power and flow. Power imposed by an impeller is proportional to the cube of its speed and the fifth power of its diameter, in liquids of low viscosity. Thus, with baffles or their equivalent, constant power can be achieved for impellers of different diameters according to the relationship, $N_r = (1/D_r)^{5/3}$, where N_r is the ratio of speed and D_r is the ratio of the corresponding impeller diameters.

Flow from impellers is proportional to the speed and to the cube of the diameter. It is clear that for the same power there will be greater flow from the large-diameter, low-speed turbine than from the small-diameter, high-speed turbine.

High-viscosity liquids and pastes usually require different techniques from those for mixing of low-viscosity liquids. Special apparatus is necessary to provide for wiping, stretching, and squeezing because turbulence cannot be generated in such fluids to provide for the small-scale mass transfer necessary to cause interpenetration of substances. There are few quantitative data yet available relating performance to the various types of equipment.

[illegible]

because classification and separation often occur after attainment of the desired distribution if the operation is carried on too long. A systematic approach to the mixing of solids and pastes has not yet been developed.

Equipment for mixing. Portable mixers generally range in size from fractional horsepower to 3 hp; they are designed for use in open tanks up to 3000-gal. capacity and are clamped to the tank shell. Direct-drive units usually run at 3600, 1750, or 1150 rpm, or at variable speed. Gear-drive units run at about 420 rpm, or at variable speed.

Light-duty, permanently mounted mixers range in size from fractional horse power to 3 hp and are used on open or closed tanks up to 3000-gal capacity.

Heavy-duty, permanently mounted, top-entering mixers are made in capacities up to 500 hp. They are normally designed for speeds between 45 and 200 rpm, are made with independent mounting for the mixer shaft, and are connected to the drive by a flexible coupling to protect the speed-reduction gearing. See UNIT OPERATIONS. [F.H.F.]

Bibliography: P. E. Graybeal and R. J. Bechtel, *Ind. Eng. Chem.*, vol. 49, March, 1957; W. L. McCabe and J. C. Smith, *Unit Operations of Chemical Engineering*, 1956; J. H. Rushton, *Fourth World Petroleum Congress, June 6–15, 1955*, sec. III/B, no. 3, 1955–1956; J. H. Rushton and J. Y. Oldshue, *Chemical Engineering Progress Symposium*, ser 55, no. 25, 1959.

Mixture

An aggregate composed of two or more distinct chemical components which retain their identities regardless of the degree to which they have become mingled. The substituents may be present in any proportion and may form a homogeneous or heterogeneous system. There is no chemical interaction between the components of a mixture, and it is possible, at least theoretically, to separate them by physical means. Attractive or repulsive forces between the different substituents, however, may affect the gross characteristics of the aggregate. For example, the total volume of a mixture of several liquids may not be the sum of the separate volumes. [F.J.J.]
See CHEMICAL COMPOUNDS.

Mizar

A multiple star in the constellation Ursa Majoris. Mizar (ζ Ursa Majoris) is located at the bend in the handle of the Big Dipper. A very wide visual double, Mizar (the horse) and Alcor (the rider), are visible to the naked eye separated by a distance of 12′. Mizar is also a telescopic double star, separation 14″; in addition, both components are also spectroscopic binary stars. Mizar was the first spectroscopic binary to be discovered, in 1889. It showed spectral lines from both stars which, because of the Doppler effect, shifted from violet to red and back, in a period of 20 days, as the two stars revolved about their common center of gravity. The amplitude and period together permit computation of the masses of the stars. Mizar is a member of the moving cluster of stars in Ursa Major. [J.L.G.]

Mockingbird

A member of the family Mimidae, *Mimus polyglottos*. This long-tailed, gray and white bird is a familiar sight in the southern United States, where



Mockingbird, *Mimus polyglottos*, length to 11 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

its song, especially the night song, is greatly admired. In recent years the mockingbird has extended its range northward and is now found in all of the United States except the northern fringe and the mountains. It is slightly migratory, but certain individuals remain the year around, even in Wisconsin and Michigan. As its name implies, it is an excellent mimic, readily adding the songs and calls of other birds to its repertoire. *See* CATBIRD; PASSERIFORMES; THRASHER. [J.D.B.]

Mode of vibration

A characteristic manner in which vibration occurs. In a freely vibrating system, oscillation is restricted to certain characteristic patterns of motion at certain characteristic frequencies; these motions are called normal modes of vibration.

An ideal string, for example, can vibrate as a whole with a characteristic frequency

$$f = (1/2L)\sqrt{T/m}$$

where L is the length of string between rigid supports, T the tension, and m the mass per unit length of the string. The displacements of different parts of the string are governed by a characteristic shape function. More specifically, the motion of any part of the string is proportional to $\sin(\pi x/L) \sin(2\pi ft)$, where x is the distance of the part of the string from a fixed end and t is the time. This simplest kind of vibration is the first, or fundamental, mode of vibration of the string; its frequency is the fundamental frequency. All parts of the string vibrate with the same frequency and all parts of the string move outward from equilibrium at the same time.

The string is also capable of vibrating in two segments, one of which goes outward from equilibrium in a positive direction at the same time as the other part goes outward in a negative direction, and conversely. Again the motion of any part of the string can be described by the product of a space function by a sinusoidal function of time: $\sin(2\pi x/L) \sin(2 \times 2\pi ft)$. All parts of the string move together as a sinusoidal function of the time and at the same frequency; the space function governs the motion in opposite directions. The frequency of this second mode of vibration is twice that of the first mode. Similarly, modes of higher order have frequencies that are integral multiples of the fundamental frequency.

Because the frequencies are in the ratios 1:2:3 the modes of vibration of an ideal string are properly called harmonics. *See* HARMONIC (PERIODIC PHENOMENA). Not all vibrating bodies have harmonic modes of vibration, however. The ideal drumhead, for example, vibrates freely with frequencies in the ratios 1:1.59:2.14:2.30. . . . In fact, most real systems vibrating freely have modes of vibration whose frequencies are not exactly in the ratios of integers. *See* VIBRATION; *see also* PARTIAL TONE. [R.W.Y.]

Model theory

A scaled model behaves similarly to its prototype. On the basis of model theory, a small, readily modified model can be built and tested at low relative cost and the results applied to the full-scale device. For the results to be applicable, similitude relationships between model and prototype, or between any other two systems that are being compared, must be satisfied.

Dimensionless characteristic. A convenient method of developing the similitude relationships makes use of the dimensionless characteristic equation for the prototype system (*see* DIMENSIONAL ANALYSIS). The characteristic equation for a system may be written as

$$\pi_1 = f(\pi_2, \pi_3, \pi_4, \ldots, \pi_s) \qquad (1)$$

in which π_1 is the dimensionless term containing the dependent variable and the other π terms are dimensionless quantities based on the independent variables. A similar equation may be written for a similar system (the model) that is usually of different size.

$$\pi_{1m} = f(\pi_{2m}, \pi_{3m}, \pi_{4m}, \ldots, \pi_{sm}) \qquad (2)$$

If the two systems involve the same phenomenon, the forms of the two functions will be identical. If the model is so designed that

$$\pi_{2m} = \pi_2, \pi_{3m} = \pi_3, \pi_{4m} = \pi_4, \ldots, \pi_{sm} = \pi_s \qquad (3)$$

it follows that

$$\pi_1 = \pi_{1m} \qquad (4)$$

Thus the dependent variable π_{1m} measured in the model may be used to predict the corresponding variable π_1 in the prototype. Equations (3) are the design equations and Eq. (4) is the prediction equation.

In general, the variables involved in fluid-me- [illegible]

systems (*see* FLUID MECHANICS; FLUID-FLOW PROPERTIES; GAS DYNAMICS). Typical forces include those arising from gravity, viscosity, compressibility, and surface-tension effects. Many dimensionless terms may be evolved for a given set of variables, but in general the terms arising in fluid-mechanics systems may be reduced to the combinations in Eq (5)

$$\pi_1 = f\left(\frac{v^2}{gd}, \frac{v}{c}, \frac{\rho v d}{\mu}, \rho\frac{v^2 d}{\sigma}, \frac{\lambda_i}{d}\right) \qquad (5)$$

in which v is velocity, g is acceleration of gravity, d is a characteristic length, ρ is the density of the fluid, μ is the viscosity of the fluid, c is the velocity of sound in the fluid, σ is the surface tension of the fluid, and λ_i is any significant length other than d. The first four quantities in the function have been named. *See* FROUDE NUMBER; MACH NUMBER; REYNOLDS NUMBER.

Thus in a fluid-mechanics system the similitude requirements from Eq. (3) are that the Froude, Mach, Reynolds, and Weber numbers in the model be equal to their counterparts in the prototype, and that the model and prototype be geometrically similar in all significant respects. If the performance of a given system is independent of one of the properties or forces, the corresponding term is omitted from the characteristic equation. In each of the dimensionless numbers the numerator is proportional to the inertia force of the fluid and the denominator is proportional to the corresponding significant force developed in the fluid. Thus equivalence of the respective numbers in model and prototype means that all forces are in the same ratios in model and prototype and hence the performances of the model and prototype will be equivalent. *See* DYNAMIC SIMILARITY. [C.M.]

Bibliography: E. A. Bonney, ***Engineering Supersonic Aerodynamics***, 1950; A. B. Cambel and B. H. Jennings, ***Gas Dynamics***, 1958; D. O. Dommasch, [illegible] 2 vols., 1953–1954.

Modulation

The process or result of the process whereby a message is changed into information-bearing signals that not only unambiguously represent the message but also are suitable for propagation over the transmitting medium to the receiver.

The vehicle for the propagation of electric signals from one region in space to another is always an electromagnetic field, and when this field changes with time it takes the form of a wave.

At the distant end the receiver is waiting to be informed; this is accomplished by the arrival of the propagated wave which, it is important to note, must change in a way the receiver cannot predict. Here too, modulation is the process whereby in response to the received wave either the original message or information pertaining to the original message is made available in the form desired and is delivered when and where it is wanted. The terms demodulation and detection are often used to denote the recovery of the wanted message from a modulated signal.

Modulation is fundamental to communication. No matter how, when, or where communication takes place, modulation is encountered (*see* COMMUNICATIONS, ELECTRICAL). Many kinds of modulation are possible, but a characteristic common to all is change.

Modulation implies bandwidth occupancy. For any signal to change in a way that cannot be predicted necessarily implies that the signal occupy a nonzero band of frequencies. For example, the spoken word occupies a band from a few hundred to several thousand cycles per second.

Ordinary telephony (Fig. 1) is a good example of these modulation concepts. Longitudinal sound waves generated by the spoken word constitute the information-bearing signals to be communicated. The telephone transmitter, acting as a modulator, changes this acoustic energy into electric energy suitable for high-speed propagation to a distant point. At the receiving end, demodulation in the telephone receiver changes the electric signals back to pressure waves in the air.

Defined broadly, modulation is the process or result of the process whereby some parameter of one wave is varied in accordance with another. As is customary in the treatment of modulation, the word "wave" is used as a generic term intended to include such concepts as signal, voltage, current, pressure, displacement, and the like, whether these are constant or changing.

This broad definition of modulation, which may be illustrated by a familiar example of amplitude modulation, implies three fundamental concepts: modulating wave, carrier, and modulated wave. As Fig. 2 clearly depicts, a modulating wave changes some parameter of the wave to be modulated; a carrier is a wave suitable for modulation by the modulating wave; and lastly, a modulated wave is a wave some parameter of which is changed in accordance with the modulating wave.

Amplitude modulation (AM). In amplitude modulation the amplitude of a carrier is the parameter subject to change by the modulating wave.

In a more restrictive sense, AM is defined to mean modulation in which the amplitude factor of a sine-wave carrier is linearly proportional to the modulating wave. Analysis shows that the modulated wave depicted in Fig. 2 is composed of the transmitted carrier, which conveys no information

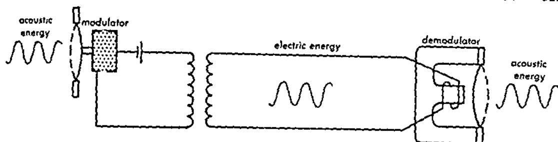


Fig. 1. Ordinary telephony. (From H. S. Black, *Modulation Theory*, Van Nostrand, 1953)

(apart from its amplitude, frequency, and phase), plus the familiar upper and lower sidebands which convey identical and therefore mutually redundant information. Thus, AM doubles bandwidth occupancy. *See* AMPLITUDE MODULATION; SIDEBAND.

Assuming adequate knowledge of the carrier, ei-ther sideband alone [illegible]

[illegible] frequency space and, when combined with frequency-division multiplexing, is widely used for long-distance telephony. For a discussion of multiplexing *see* TRANSMISSION THEORY AND METHODS.

Angle modulation. Instead of conserving bandwidth, bandwidth occupancy may be intentionally increased in exchange for improved performance. Angle modulation is one of the simplest ways of sacrificing frequency space for reduced noise. In angle modulation the angle (entire argument) of a sine-wave carrier is the parameter changed by the modulating wave. Frequency and phase modulation are particular forms of angle modulation. Often the term frequency modulation is used to connote angle modulation. *See* ANGLE MODULATION.

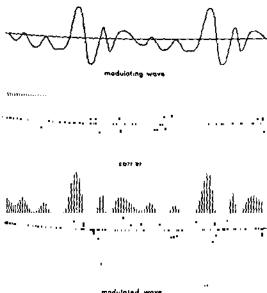


Fig. 2. Amplitude modulation of a sine-wave carrier by a voice-frequency modulating wave; typical waveforms. (From H. S. Black, *Modulation Theory*, Van Nostrand, 1953)

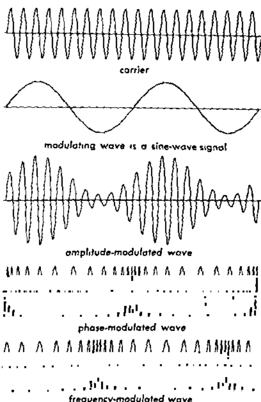


Fig. 3 Amplitude, phase, and frequency modulation of a sine-wave carrier by a sine-wave signal. (From H. S Black, *Modulation Theory*, Van Nostrand, 1953)

A fundamental concept in angle modulation is that of instantaneous frequency. It is proportional to the time rate of change of the angle of a sine function, the argument of which is a function of time. When the argument is expressed in radians and time in seconds,

$$\text{Instantaneous frequency} = \frac{1}{2\pi}\frac{d}{dt}\quad(\text{angle})$$

Frequency modulation (FM). FM is angle modulation in which the instantaneous frequency of a sine-wave carrier is caused to depart from the carrier frequency by an amount proportional to the instantaneous value of the modulating wave (see Fig. 3). *See* FREQUENCY MODULATION.

Phase modulation (PM). PM is angle modulation in which the angle of a sine-wave carrier i-

Principal advantages, disadvantages, and uses of typical kinds of modulation

Kind	Advantages	Disadvantages	Uses
Amplitude modulation			
Double-sideband plus carrier (AM)	Simplifies receiver; preserves waveform of wanted message	Doubles bandwidth occupancy; requires extra signal power	Radio broadcasting, TV broadcasting, telephony, telegraphy, telemetering
Single-sideband, suppressed carrier (SSB)	Saves bandwidth occupancy; conserves signal power	Unable to handle relatively low frequencies; adds inherent delay; waveform of wanted message is not preserved	Long-distance telephony and telegraphy over land and submarine cables
Vestigial sideband (VSB)	Avoids disadvantages of single sideband	Modest increase in bandwidth occupancy as compared with single sideband	TV broadcasting
Day's system (special phase discrimination multiplexing applicable to pairs of channels)	Conserves bandwidth and signal power	Sensitive to transmission impairments	To multiplex the two so-called color components in color TV broadcasting
Phase discrimination multiplexing (applicable to many channels)	Conserves bandwidth and signal power	Sensitive to transmission impairments	Minor
Angle modulation			
Narrow-band	Constant signal power	Extra bandwidth occupancy	Telecommunications, particularly broad-band carrier and TV over microwave radio relay systems
Wide-band	Reduces noise in exchange for extra bandwidth occupancy, constant signal power, channel-grabbing property	Extravagant of bandwidth occupancy; sensitive to some forms of transmission impairment; signal power must be adequate to override wide-band noise	Telecommunications generally, including such fields as telegraphy, telephony, radio broadcasting, telemetering, mobile communications for military and peacetime services, navigational aids, maritime beacons, meteorological aids
Wide-band with negative feedback receiver	Reduces noise in exchange for extra bandwidth occupancy; constant signal power; stabilizes receiver and improves its linearity; signal power need only override noise in a narrow band directly above and below the carrier	Extravagant of bandwidth occupancy; sensitive to some forms of transmission impairment	Negligible; suitable where the noise advantage of wide-band angle modulation is important and signal power must be conserved
Pulse modulation			
Amplitude (PAM)	Permits multiplexing channels by time division	Sensitive to some forms of transmission impairment	Radio, radar, telegraphy, telephony, telemetering time-multiplexed sampled-data systems, computers, switching
Duration (PDM)	Permits multiplexing channels by time division; reduces noise in exchange for extra bandwidth occupancy; constant signal power	Extra bandwidth occupancy; pulses vary in duration	Microwave radio relay systems, telemetering
Position (PPM)	Permits multiplexing channels by time division; reduces noise in exchange for extra bandwidth occupancy; constant signal power; saves signal power as compared with PDM	Extra bandwidth occupancy; pulses vary in position	Microwave radio relay systems, telemetering

Principal advantages, disadvantages, and uses of typical kinds of modulation (Cont.)

Kind	Advantages	Disadvantages	Uses
Code (PCM)	Permits multiplexing channels by time division; in exchange for extra bandwidth occupancy, tolerates considerable noise and serious transmission impairments, and may be repeatered again and again without significant distortion, constant signal power	Extra bandwidth occupancy, transmits digital instead of analog signals, thereby introducing quantization noise if the receiver delivers analog signals	Multiplex telephony and telegraphy, TV, data processing, combined transmission and switching systems, telemetering
Multiple modulation	Many, depending upon circumstances; often accomplishes what cannot be done in one step	Extra steps usually add extra complexity	A feature of all but the simplest transmitters and receivers

caused to depart from the carrier angle by an amount proportional to the instantaneous value of the modulating wave. The instantaneous frequency deviates from its unmodulated value by an amount proportional to the time derivative of the modulating wave (see Fig. 3). See PHASE MODULATION.

FM and PM are similar in the sense that any attempt to shift frequency or phase is accomplished by a change in the other. The terms FM and PM simply indicate which parameters of the complete argument (angle) of the sine function are being modulated.

Pulse modulation. In pulse modulation the carrier may be a train of regularly recurrent pulses. Modulation may control the amplitude, duration, position, or mere presence of the pulses so as to represent the message to be communicated. These forms of pulse modulation are commonly called, respectively, pulse-amplitude modulation, pulse-duration modulation, pulse-position modulation, and pulse-code modulation. Ease of multiplexing channels by time division is one of the important economic advantages of all forms of pulse modulation. All but pulse-amplitude modulation can exchange extra bandwidth occupancy for noise reduction. Pulse-code modulation is a digital system transmitting ON or OFF pulses and thereby offering major transmission advantages not possible with an analog system. See PULSE MODULATION.

Multiple modulation. Practical applications of modulation often utilize what is commonly referred to as multiple modulation, a succession of modulating processes in which the modulated wave from one process becomes the modulating wave for the next. A typical example is a system in which position-modulated pulses are used to modulate the amplitude of a sine-wave carrier. Such a system is abbreviated PPM-AM (pulse-position modulation–amplitude modulation) inasmuch as it is customary to list the processes in the order in which the message to be conveyed encounters them. A subcarrier is any carrier used in an intermediate step of multiple modulation.

Engineering applications. The principal application and major fields of use of these various kinds of modulation are indicated by the technical comparisons and typical examples in the table.

[H.S.BL.]

Bibliography: H. S. Black, *Modulation Theory*, 1953. M. Schwartz, *Information Transmission, Modulation, and Noise*, 1959.

Modulator

Any device or circuit by means of which the desired signal is impressed upon a periodic radio-frequency wave, called the carrier. The process is called modulation. For a discussion of it see MODULATION.

Transmission of a signal can be effected by the variation of the amplitude, the frequency, or the phase of the carrier. The methods for accomplishing this and the practical circuits will be found to differ in each case.

Amplitude modulator. In amplitude modulation the envelope of the radio-frequency carrier must be varied in accordance with the signal being transmitted. There are many ways to accomplish this, but in all cases it is necessary to employ a nonlinear device whose characteristic behavior at the carrier frequency can be made to vary at the signal frequency. For example, in absorption modulation the radio-frequency power passing through a network is partially absorbed by a series or shunt resistor whose magnitude is made to vary in proportion to the signal voltage. Alternatively, the radio-frequency carrier can be passed through an amplifier whose radio-frequency gain is varied by altering the applied potentials at signal frequency. See AMPLITUDE MODULATOR.

Frequency modulator. In frequency modulation, the instantaneous frequency of the carrier is varied in accordance with the signal frequency, while the amplitude of the carrier remains constant. One method is to vary the frequency of an oscillator by placing into its frequency-determin-

ing circuit a reactance whose magnitude can be made to vary with signal frequency. Alternatively, the phase of a radio-frequency carrier passing through an amplifier can be changed in accordance with the signal frequency. The resultant phase-modulated carrier can be altered into a frequency-modulated wave by means of subsequent stages of frequency multiplication. *See* FREQUENCY MODULATOR.

Phase modulator. In phase modulation, the signal can be transmitted by varying the phase of the transmitted wave. The difference between phase modulation and frequency modulation is principally one of definition, and most circuits suitable for the generation of frequency-modulated waves can be altered to provide phase modulation as well. *See* PHASE MODULATOR. [E.L.G.]

Mohair

The long, lustrous hair fiber of the Angora goat, native to the province of Angora, Turkey. This species of goat is now raised also in the United States, principally in the Southwest. The United States, Turkey, and the Union of South Africa are the most important producers of mohair (Table 1).

According to the Crop Reporting Board (AMS, USDA) in a 1958 release there were 2,865,000 angora goats in Texas, valued at $25,800,000 representing about 97% of the industry in the United States (Table 2). Imported mohair is of long staple, 9–12 in. long, and represents 1 full year's growth. The domestic goat is shorn twice a year, yielding a shorter staple, 8–10 in. long (Table 3). Imported mohair can be spun to a fineness of 60s in yarn count. The highest count possible for domestic fiber is 40s (Table 4). The domestic fiber contains a large amount of coarse, stiff hair, known as kemp, which does not process readily or allow thorough penetration of [illegible]

re[illegible]

it enhances color, beauty, softness, and luster in fabrics, and it does not attract or hold dirt particles. Mohair absorbs dye evenly and permanently and permits unusual decorative effects. The fiber is more uniform in diameter than wool (*see* WOOL).

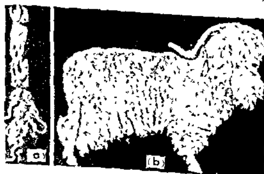


Fig 1. (a) Ringlet lock of mohair. (b) Ringlet type of Angora buck. (*M. Camp, Texas Sheep and Goat Raisers' Magazine*)

Table 1. Production of mohair, grease basis, in Turkey, Union of South Africa, and United States, 1950–1954, inclusive*

Year	Turkey, million lb	Union of South Africa, million lb	United States, million lb
1950	12.9	5 8	13 2
1951	14.9	5.4	12 9
1952	16.4	5 4	12 1
1953	16 5	5 5	12 6
1954	16.6	5.8	13 7

* From *Supplement to Wool Statistics and Related Data* USDA Statist. Bull 142, 1955.

Table 2. Production of mohair in the seven principal producing states, 1950–1954, inclusive*

Yearly production, 1000 lb	1950	1951	1952	1953	1954	Total
Texas	12,643	12,280	11,561	11,972	13,097	61 553
New Mexico	176	155	163	150	152	796
Oregon	162	152	130	152	114	710
Missouri	130	161	108	140	144	683
Arizona	88	93	109	112	120	522
California	33	34	33	30	30	160
Utah	13	13	12	16	16	70
Total	13,245	12,888	12,116	12,572	13 673	64 494

* From *Supplement to Wool Statistics and Related Data* USDA Statist Bull 142, 1955

Table 3. Mohair: number of goats clipped in the United States, 1956 and 1957*

Year	Number goats clipped†	Average clip per goat, lb	Total production, lb	Price per pound, cents	Total value of sales
1956	3,164,000	5 8	18,233,000	84 4	$15 383 000
1957	3,246,000	5 9	19,072,000	88 6	16 895,000

* From USDA Agricultural Marketing Service, Statist. Bull 230, July, 1958

† In states where goats are clipped twice a year, the number clipped is the sum of goats and kids clipped in spring and kids clipped in fall

Table 4. Tentative specifications for fineness of mohair tops*

Grade	Average diameter of fiber, microns	
	Minimum	Maximum
40s	23 55	25 54
36s	25.55	27.54
32s	27.55	29 54
30s	29.55	31 54
28s	31.55	33 54
26s	33 55	35 54
22s	38 05	40 54
18s	38.05	40 54

* American Society for Testing Materials, *Standards on Textile Materials*, January, 1956

Its strength and freedom from felting make mohair particularly valuable.

Uses. When mohair is used in pile fabrics, the naturally strong fiber, combined with the strength of the pile weave, makes an especially durable and serviceable fabric. Mohair fabrics are wrinkle resistant and do not mat readily because of the natural resiliency of the fiber. The fabric can be

made mothproof. Because mohair is resilient and stronger than wool or the other hair fibers, it is used to great advantage in rugs and quality floor coverings, and in better grades of upholstery and drapery materials. It lends itself well to embossing and hand-block printing. Other uses include men's suiting, ladies' dress and coat material, imitation furs, and wigs for theatrical purposes.

Mohair technology. This concerns the structure, growth characteristics, and chemical properties of mohair affecting and determining commercial uses.

Physical properties. Average fleece yield per goat in 1954 was 5.5 lb. The primary fleece types are the ringlet, the intermediate, and the flat lock. Ringlet types are associated more with fleeces of extreme fineness than the other types (Fig. 1). The flat lock type is usually associated with a heavy shearing weight, and although not valued as highly as the ringlet type, is a desirable quality of hair (Fig. 2). Open or fluffy fleeces are objectionable. Most breeders are endeavoring to establish an intermediate fleece between the ringlet and the flat lock types (Fig. 3). Inheritance of these traits has not been established.

Grading. Methods for grading wool and mohair are similar. Fineness is based upon mean fiber diameter measured in microns.

Structure. Microscopic structures of wool and mohair are similar. Epidermal scales of mohair, however, are tight, faintly visible, hardly overlapping (Fig. 4). This gives mohair high luster, smooth handling qualities, but lack of felting ability. The cortical layer is composed of spindle cells which surround a medulla.



Fig. 2. (a) Flat lock of mohair. (b) Flat lock type of Angora buck. (M. Camp, Texas Sheep and Goat Raisers' Magazine)

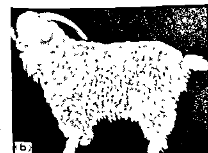


Fig. 3. (a) Intermediate lock of mohair. (b) Intermediate type of Angora buck. (M. Camp, Texas Sheep and Goat Raisers' Magazine)

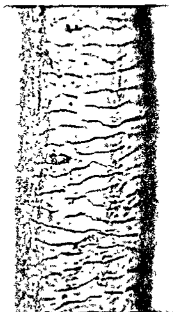


Fig. 4. Scale impression of mohair fiber. Scales are tight, hardly overlapping.

Factors affecting mohair. Several factors influence the quality and quantity of mohair. (1) Sex: this determines quantity in descending order as follows, bucks, wethers, dry does, does, kids. (2) Age: weight of fleece increases with age up to 5 years. The quality changes drastically with age; kid hair is fine and adult hair coarse. (3) Season of shearing: fall-sheared fleeces are lighter and finer than spring-sheared fleeces. (4) Condition of the goat: sparse, droughty ranges result in reduced fleece weights and fiber diameter. (5) Type of lock formation: whether ringlet, intermediate, or flat, also influences the quality of mohair. *See* ALPACA, CAMEL'S HAIR; CASHMERE; LLAMA; *see also* FIBER, NATURAL. [T.D.W.]

Bibliography: See AGRICULTURAL SCIENCE (ANIMAL).

Moho (Mohorovicic discontinuity)

A seismic discontinuity at the base of the earth's crust assumed as an explanation of travel time curves indicating that seismic waves undergo a sudden increase in velocity. Immediately below the discontinuity the velocity of seismic compressional waves increases to a little over 8 km/sec. The discontinuity is named after its discoverer, the Yugoslavian geophysicist A. Mohorovičić, who noticed the sudden change in velocity when examining earthquake records of the Yugoslavian Kulpa Valley earthquake of October 8, 1909. Such a discontinuity seems to appear every place that geophysicists seismologically investigate the earth's structure. Under the ocean basins its depth is generally 10–12 km; under the continents it is usually at a depth of 33–35 km (Fig. 1). It usually

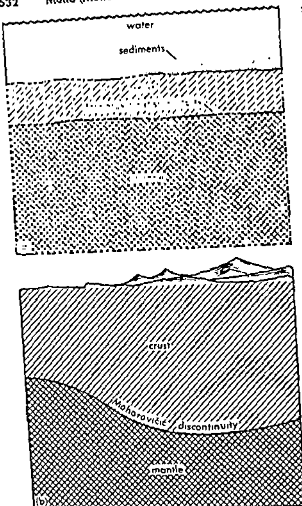


Fig. 1. Diagrammatic portrayal of the Mohorovičić discontinuity (a) At average depths some 10–12 km beneath ocean basins. (b) At average subcontinental depths of 33–35 km.

lies deeper under mountain masses and shallower under lowlands and plains, sloping upward near coastal margins (*see* CONTINENT). Islands and island arcs are also marked by a downward bend of the Moho under the ocean basins. *See* SEISMOLOGY.

Rock density. In addition to the increase in velocity there is also thought to be an increase in the density of the rocks as they change from crustal to mantle material. The density of the crust varies from 2.2 to 2.7 g/cm³ and the density in the surface of the mantle is 3.3. The crust is composed of light, silicic rocks which surround, or ride upon, the denser, basic mantle.

Composition. Although the Moho is generally accepted as the dividing zone between the crust and mantle, there is still considerable speculation over its composition and configuration. M. Ewing has noted that in some localities a peculiar scattering takes place in the seismic waves which would indicate a rough surface, perhaps the remnants of an [illegible]

[illegible] the Moho represents, at least partly, the beginning of the mantle. [illegible] its composition is in question by some and accepted as peridotite, eclogite, dunite, or serpentine by others. Those who accept the composition of the outer mantle base their reasoning upon laboratory seismic velocity measurements, theory, and in one instance, the outcropping of St. Paul's Rocks in the middle Atlantic. At St. Paul's Rocks peridotite is found above sea level along a tremendous vertical fault. It has been postulated by H. Hess that the crust is broken through and heaved up along one side of the fault to the extent that the mantle is exposed. In spite of these theories, the nature of the Moho and outer mantle will not be determined completely until one or more core holes are drilled through the crust and samples are collected.

In certain areas of the ocean basins the Moho is relatively close to the surface (sea level). In the area just north of San Juan, Puerto Rico, on the rise to the north of the Puerto Rican Trench, the Moho has been observed, by J. Nafe, to lie at a depth of 9+ km. At the southern end of the Gulf of California it has been observed, by G. Shor, at a depth of 9+ km. It would be possible to reach the Moho by drilling in either of these locations. Contemplation of and tentative planning for such drilling brought about the coining of another word, Mohole.

Temperature and pressure. The question as to whether presence of the Moho is evidence of chemical changes in the rocks of the crust and mantle or only of a physical phase change is still unanswered. Recent evidence is cited in favor of the phase change, suggesting that the Moho results from conditions of temperature and pressure which cause the transformation of a basalt crust into an eclogite mantle. Using the average for crustal

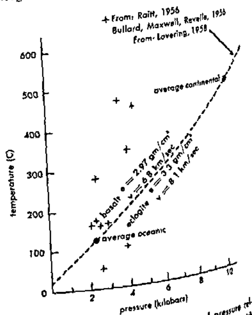


Fig. 2. Relationship of temperature and pressure relative to the Moho as suggested by T. S. Lovering with seismic and geothermal data from the Pacific recently reported by R. W. Raitt and by E. C. Bullard et al.

Table of Moho depths*

Location	Seismic velocity, km/sec	Depth to Moho, km
East coast, U.S.	8.10	33
New York–Pennsylvania	8.15	35
Central-Appalachian	8.03	40
Canada	8.10	36
Wisconsin	8.10	40
Southern California valleys	8.10	32
South Africa	8.20	36
Deep Atlantic (aver.)	8.10	10
Pacific Basin (aver.)	8.20	11

*After B. Gutenberg.

thicknesses beneath the continents and oceans and assuming that the heat flow is similar in the two locations, a relation has been derived, illustrated by the dashed line in Fig. 2. However, recent seismic and geothermal results from the Pacific, which are also shown in Fig. 2, tend to controvert this idea. They indicate temperatures at the Moho that are far in excess of those which have been assumed.

Depth summary. The table presents a summation of some continental and oceanic seismic velocities and Moho depths from refraction measurements. [C.C.L.]

Bibliography: B. Gutenberg, Wave velocities in the earth's crust, *Crust of the Earth*, Geol. Soc. Am. Spec. Paper 62, 1955; H. Jeffreys, *The Earth: Its Origin, History and Physical Constitution*, 4th ed., 1959; H. E. Landsberg (ed.), *Advances in Geophysics*, vol. 3, 1956; G. G. Lill and A. E. Maxwell, The earth's mantle, *Science*, 129(3360):1407–1410, 1959.

Mohole

A term coined by scientists and others to designate plans for, and interest in, drilling a bore hole into and through the Mohorovičić discontinuity. This discontinuity, or the so-called moho, is generally considered to be the dividing zone between the crust and the mantle rock zones of the earth. Many consider that present seismic knowledge of thinner crust beneath the oceans and current development of submarine drilling for petroleum make such a subterranean project feasible. See EARTH; MOHO (MOHOROVICIC DISCONTINUITY); SEISMOLOGY. [C.V.C.]

Moisture-content measurement

Measurement of the ratio or percentage of water present in a gas, a liquid, or a solid (granular or powdered) material. Nearly all materials contain free water, the relative amount being dependent upon the physical and chemical properties of the material. The primary purpose of determining and maintaining moisture contents within specified limits can usually be traced to economic factors, trade practices, or legal requirements.

Moisture content has a number of synonymous terms, many of which are specific to certain industries, types of product, or material. The water content in solid, granular, or liquid materials is usually referred to as moisture content on either the wet or dry basis; the wet basis is common to most industries. Specifically, moisture content on the wet basis refers to the quantity of water per unit weight or volume of the wet material. A weight basis is preferred. The textile industry uses the dry basis for moisture content of textile fibers. Often referred to as regain moisture content, the dry basis or regain refers to the quantity of water in a material expressed as a percentage of the weight of the bone-dry (thoroughly dried) material. The relationship between the wet and dry moisture-content basis is shown in Fig. 1.

The moisture content in air is referred to as humidity, either absolute or relative. Absolute humidity is the number of pounds of water vapor associated with 1 lb of dry air, also called just humidity. Relative humidity is the ratio, usually expressed as a percentage, of the partial pressure of water vapor in the actual atmosphere to the vapor pressure of water at the prevailing temperature. Relative humidity (RH) is customarily reported by the U.S. Weather Bureau because it [illegible]

venient for many purposes such as computations used in air-conditioning, combustion, or chemical processing; therefore absolute units, such as dew-point or grains of water per pound of dry air, are more acceptable. Dew point is the temperature at which a given mixture of air and water vapor is saturated with water vapor.

MOISTURE CONTENT OF GASES

The measurement of water content in gases and mixtures of air and gases is important in industry. A number of commercially manufactured instruments is available for these measurements; their principles of operation include condensation, used in dew-point or fog-point indicators; dimensional change, used by hygrometers; thermodynamic equilibrium, used by wet-bulb psychrometers; and absorption methods, which serve as the basic prin-

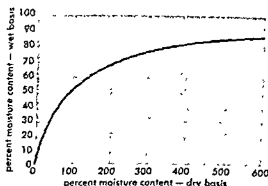


Fig. 1. Relation of dry vs. wet moisture-content basis

ciple for gravimetric and electric conductivity or dielectric types.

The importance of humidity in relation to personal comfort is well known. As a result the air-conditioning industry has grown to considerable proportions by producing equipment to maintain comfortable conditions of temperature and humidity. Considerable industrial air-conditioning is also done for process reasons. The textile industry makes wide use of humidity control in rooms for weaving, carding, spinning, and other processes, because the amount of moisture held or absorbed by the textile fibers affects these operations. Paper manufacturers face problems similar to those of the textile industry, as do certain chemical, plastic, and allied processing industries where control of humidity is important for product quality.

Control of humidity is important in the preservation of materials, especially those which are hygroscopic, and in the storage of food products. In many instances, humidities must be maintained at high levels, as in the storage of apples and vegetables, whereas humidity is maintained at low levels in the storage of dried milks, eggs, and similar products.

Human comfort is affected by high humidities, because the air is so close to its saturation content that it cannot absorb moisture from the surface of the skin, and thus cool the individual by evaporation. The higher the temperature of the air, the greater the amount of moisture it can hold. A rule of thumb for human discomfort is that any combination of temperature and relative humidity totaling 130 or higher is uncomfortable. *See* HUMIDITY; HUMIDITY CONTROL.

Psychrometers. A psychrometer is a device for measuring moisture content of air or gases by means of two thermometers. One thermometer bulb is covered with a wick and maintained wet (the wet bulb); the other bulb is exposed directly to the air or gas (the dry bulb). The evaporation of water from the moistened wick of the wet bulb produces a lowering of its temperature, and by observation of the difference in temperature between the two bulbs the absolute or relative humidity may be determined. For accurate results, the gas or air must have a velocity of 15–20 ft/sec past the wet bulb. Psychrometric charts or tables are used with the readings obtained from the two thermometers to determine the moisture content of the air or gas. Instruments utilizing this principle are often called wet- and dry-bulb thermometers *See* PSYCHROMETER.

Hygrometers. Hygrometers measure humidity by the change in dimensions of a hygroscopic material, such as human hair, organic membranes, wood, and plastics. Their most dependable range of operation is from 15 or 20% to 85 or 90% relative humidity at temperatures of 0° to approximately 160°F. Stability of these instruments is better when they are not subjected to extremes of temperature or humidity. There is considerable time lag in the system response to changing humidity conditions

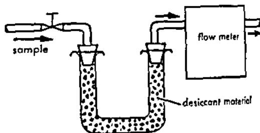


Fig. 2. Gravimetric absorption method.

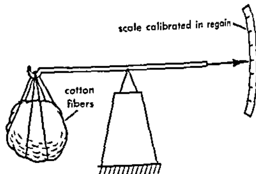


Fig. 3. Aldrich regain indicator.

when operated at low temperatures. Accuracies of ±3% relative humidity can be expected at normal room temperatures.

Another type of hygrometer utilizes changes in the electrical conductivity of a hygroscopic salt as its operating principle. The conductivity of a solution is dependent upon its concentration (amount of water if the salt content is constant) and its temperature. *See* HYGROMETER.

Gravimetric hygrometery. The change in weight of an absorbing material can be used to measure its moisture content. The measurement of moisture content in gases is obtained by passing a known volume of the gas through a suitable desiccant, such as phosphorus pentoxide, silica gel, or similar material, and observing its change in weight. This method is considered a primary standard and is often used in the exact calibration of instruments. It is necessary to make certain that all of the gas or atmosphere has been in contact with the desiccant for a sufficient time to ensure complete absorption of the water vapor. Exact initial and final weighings are also necessary. Figure 2 shows the basic principle of this instrument.

Hygroscopic materials may also be employed directly to determine changes in water vapor in air. The Aldrich regain indicator makes use of a loose ball of cotton attached on one arm of a sensitive balance; the other arm of the balance serves as a pointer as in Fig. 3. It is important that the sensing material change weight only with change in relative humidity of the surrounding atmosphere.

Dew-point indicators and recorders. When water vapor is cooled, a temperature is reached at which the phase changes to a liquid or solid. This temperature is known as the dew point. The classi-

cal method of determining the dew point consists of slowly cooling a polished metal surface until condensation takes place; the temperature at which the first droplet appears is taken as the dew point. The manually operated dew cup is one of the simplest forms of dew-point apparatus. This consists of a polished metal cup containing ether or another volatile liquid into which is placed an accurate mercury-in-glass thermometer. Air is bubbled through the volatile liquid, lowering its temperature and the temperature of the metal cup until dew (condensation) forms on the cup (see Fig. 4).

These instruments are capable of high accuracy, provided correct techniques are employed. This technique is widely used for measuring water vapor in flue gases, gasoline vapors, furnace gases, compressed gases, and others. More refined instruments utilize refrigeration and closed systems to contain the sample while the condensation is viewed through inspection windows. Others automatically record the condensation point by means of photocells, which alternately control heat or refrigeration to the target area. An adaption of the dew-point technique makes use of the Wilson cloud chamber principle. The gas sample is compressed and then vented to atmosphere, producing cooling by adiabatic expansion. By repeated trials a pressure ratio can be obtained at which a cloud or fog forms. Dew point is computed from the ratio of the initial to final pressure.

Miscellaneous methods. Several alternate methods exist for determining the amount of water vapor present in gases or air.

The difference in thermal conductivity of dry air and air with water vapor may be determined by measuring the difference in the electrical resistance (temperature) of a hot wire sealed in a small cell. This method is affected by changes in gas composition. Usually a bridge circuit is used, with a hot-wire cell containing dry air as the reference and a cell with the sample to be tested as the unknown. See BRIDGE CIRCUIT.

Spectroscopic methods and index of refraction have been used experimentally, as well as the

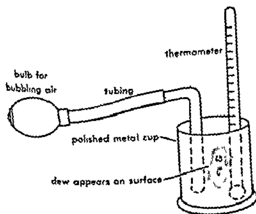


Fig. 4. Manually operated dew-point unit.

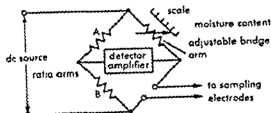


Fig. 5 Basic bridge circuit for conductivity method of moisture-content measurement.

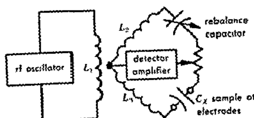


Fig 6 Basic bridge circuit for capacitance method of moisture-content measurement

measurement of pressure or volume after the absorption of the water vapor from a sample.

MOISTURE CONTENT OF LIQUIDS AND SOLIDS

The measurement of moisture in solid or liquid materials can be determined by various physical and electrical methods.

Electrical conductivity methods. These methods are based on the relationship between dc resistance and moisture content for such materials as wood, textiles, paper, grain, and similar products. Specific resistance plotted against moisture content results in an approximate straight line up to the moisture saturation point. Beyond the saturation point, where all of the cells and intermediate spaces are saturated with free water, conductivity methods are not reliable. This point varies from approximately 12 to 25% moisture content, depending on the type of product.

The sample under test is applied to suitable electrodes in the form of needle points for penetrating into woods, plaster, and similar products or flat plates for sheet materials. Granular or fibrous materials may make use of electrodes in the form of a cup or clamp arrangement to confine the material to a fixed volume. The electrodes and sample under test comprise one arm of a Wheatstone bridge as shown in Fig 5. The high sensitivity required of the detector dictates the use of electronic amplifiers. The range of resistance values corresponding to normal moisture contents varies from less than 1 megohm to 10,000 megohms or higher, depending upon material, electrode design, and moisture content. Increasing moisture content results in decreasing resistance values.

Electrical capacitance methods. These methods are based on the principle of the change in dielectric constant between dry and moist conditions of a material. The dielectric constant of most

vegetable organic materials is 2–5 when dry. Water has a dielectric constant of 80; therefore the addition of small amounts of moisture to these materials causes a considerable increase in the dielectric constant. The material being measured forms part of a capacitance bridge circuit, which has rf power applied from an electronic oscillator (Fig. 6). Electronic detectors measure bridge unbalance or frequency change, depending on the method employed

Electrode design varies with the type of material under test. Parallel-plate types are used for sheet materials, whereas cylindrical electrodes are usually adapted for liquids or powders. Modifications of the parallel-plate capacitor into a rectangular enclosure are used with granular, fibrous, or powdered materials. Range of moisture contents measurable is from 2 or 3% to 15 or 20%, varying with the product.

Equilibrium methods. The equilibrium moisture content of the air at the surface of a material is representative of the moisture within the material. This is particularly true of hygroscopic granular or fibrous materials. Humidity measuring instruments are employed by inserting the measuring element into the material. Moisture content is determined from data of relative humidity vs. moisture content for the material under test.

Absorption methods. The water content of liquid organic compounds can be determined by the use of spectrophotometers. The sample is placed in a suitable cell and monochromatic infrared radiation is passed through it. Water vapor absorbs the infrared radiation, but the material absorbs little radiant energy. Thus the radiation varies with the water content. The radiation is focused on a sensitive thermocouple and the resulting voltage amplified The indication of a sensitive galvanometer can be calibrated in terms of moisture content. This method has been used for accurately detecting 0–10 parts per million of water in Freon-12.

Chemical methods. A widely used chemical titration method is the Karl Fischer technique. The method offers extreme sensitivity with good accuracy and covers a wide range of materials. Small samples can be analyzed readily. End points can be determined either visually or electrometrically. The Karl Fischer reagent is added in small increments to a glass flask containing the sample until the color changes from yellow to brown or a change in potential is observed at the end point. Dark-colored solutions require electrical end-point measurement. This method does not differentiate between free and combined water.

Various other methods involve chemical reactions with the free water in liquids or solids by (1) the evolution of free acidic or basic compounds, (2) the evolution of an inert gas, or (3) the formations of an insoluble precipitate. In general, these methods fail to distinguish active hydrogen from water in many materials and are reliable only under laboratory control.

Distillation methods. A representative sample of the material can be placed in a suitable flask with an excess of liquid, which usually has a boiling point higher than water but is immiscible with water. Heat is applied to the flask, and the water and some of the liquid are distilled off. The combined vapors are condensed and collected. The water is separated and measured volumetrically to calculate the moisture content. A variety of laboratory glass apparatus is available for distillation of liquids which are either lighter or heavier than water. The measuring tubes are etched with graduations in cubic centimeters corresponding to grams of water. The moisture content is calculated knowing the weight of the original sample and the weight of water. This method is slow, and the accuracy depends upon the care exercised and the apparatus used. The method was widely used in the analysis of grains but has largely given way to electrical or oven drying methods.

Oven drying methods. The oldest and most common analytical method of determining moisture consists of heating the sample to ensure complete drying. Moisture is calculated on the basis of loss in weight between original and dried sample. It is applicable to many solids and some liquids and does not require unusual operator skill. Semi automatic drying and weighing ovens are available and the moisture content is indicated directly by the weighing scale built into the oven. Problems involve making certain that the material has not lost some of its weight by loss of volatile products other than water, that the material is completely dry at the time of final weighing, and that accurate weighings are made. The oven method is widely used and often serves as the primary standard for calibration of electrical and other indirect methods.
[L.E.C.]

Molar volume

The volume occupied by 1 gram-molecular weight (1 mole) of a substance, whether in the form of a solid, a liquid, or a gas. The term *molar volume* is often applied to the normal molar volume of an ideal gas, that is, to the volume occupied by 1 gram mole at 0°C, under a pressure of 1 standard atm

For an ideal gas the perfect gas law gives

$$pV = \frac{mRT_0}{M} \tag{1}$$

where p is the pressure, V is the volume occupied, m is the mass of the gas in grams, M is the gram-molecular weight, T_0 is the absolute temperature at the melting point of ice, and R is the gas constant. If $m/M = 1$ and $p = 1$, then

$$V = V_0 = RT_0 \tag{2}$$

where V_0 is the normal molar volume in the units chosen. This is an important constant of physical chemistry and it is determined by the method of limiting densities. A real gas satisfies Eq. (1) only

at zero pressure. Hence for a real gas

$$(pV)_0 = \frac{mV_0}{M} \quad (3)$$

where $(pV)_0$ is the value of pV at zero pressure. If a quantity $L = m/(pV)$ is defined, then $L_1 = m/(pV)_1$ and $L_{lim} = m/(pV)_0$. Since $m/V = \rho$, the density $L = \rho/p$ and $L_1 = \rho$ for $p = 1$. Thus L_1 is the actual density at 1 atm, and L_{lim} is the density the gas would have, at 1 atm, if it were an ideal gas. Both L and pV vary with p for a real gas and L_{lim} and $(pV)_0$ are the limiting values (obtained by extrapolation) at $p = 0$ of ρ/p

$$V_0 = \frac{M}{m}(pV)_0 = \frac{M}{L}$$

and hence V_0 is found in this way by extrapolating the density of a real gas divided by its pressure to zero pressure. The value obtained is $V_0 = (22.4146 \pm 0.0006) \times 10^3$ ml.

The molar volumes of organic liquids, measured at their boiling points are to a great extent additive properties of the liquids, and from them, the volume equivalents of the elements can be calculated. Thus, in a homologous series of hydrocarbons or hydrocarbon derivatives, there is a constant difference of 22 ml in the molar volume for each $—CH_2—$ group. The volume equivalent of two hydrogens is estimated by subtracting $n \times 22$ for n $—CH_2—$ groups from the molar volume of a paraffin C_nH_{2n+2}; the mean value is 11.0, so that each hydrogen atom contributes 5.5 ml, and hence, 11.0 ml must be the volume equivalent of a carbon atom. Other volume equivalents can be found in a similar way. *See* GAS; MOLECULAR WEIGHT; PARACHOR; SOLUTION. [T.C.W.]

Mole

Any of several small, fossorial mammals of the family Talpidae, found in North America and Eurasia. Moles are stout-bodied and highly modified for digging, with the forelimbs short and strong, the forepaws flattened, and the palms turned outward. Their fur is short, soft and silky, usually gray or bluish-gray, but varying from light brown to black; they lack color pattern variation. The eyes are vestigial and the external ear is absent.

Most moles eat earthworms and insect grubs found as they burrow just beneath the surface. The location of their burrows is marked by a low ridge of earth. They also deposit small mounds of loose soil at irregular intervals along their trails. *See* INSECTIVORA. [J.D.B.]



Common mole, *Scalopus aquaticus*; length, mole 7½ in. (From E. L. Palmer, *Fieldbook of Natural* ... McGraw-Hill, 1949)

Mole (skin)

A local growth of the skin which is usually rounded, raised, and pigmented. It is related to birthmarks and is correctly called a nevus. Nevi may be present at birth or may develop later. They are almost universally present but vary greatly in size, color, and distribution, as well as in components. The type of cell involved largely determines the external appearance of the mole and whether it will contain hair, oil glands, or special structures. Color ranges from light brown to greyish blue. Some nevi are flat, others are elevated, and some are berry-shaped.

All nevi may become malignant but the tendency is more marked in the flat, smooth, hairless, darkly colored type, especially when found on the lower extremities. Conversely, the light brown, hairy mole present from birth is both the most common form and the least likely to show malignant change.

Most moles are soft, but some are corrugated and warty. Those present at birth represent a form of birthmark, cellular in nature, in contrast to vascular birthmarks, such as port-wine stains.

Mechanical irritation, trauma, and inadequate removal may predispose to malignant changes. Any mole which shows a change in color, size, or hardness should be referred to a physician.

Finally, suspect moles must be carefully distinguished from their less common but malignant counterparts, the junctional nevi and the melanomas. [E.C.ST.]

Molecular association

The formation of double molecules or polymolecules from single species as a result of specific and moderately strong intermolecular forces. The phenomenon is encountered in the solid, liquid, and gaseous states, and is usually restricted to cases describable by equilibrium theory. Its importance arises from the insight which it provides into the structure of molecules and the nature of the interatomic forces.

Hydrogen bonding. One kind of molecular association, important for its widespread occurrence, is hydrogen bonding. In molecules in which hydrogen atoms are attached to small electronegative atoms, such as nitrogen, oxygen, or fluorine, the hydrogen atoms may act as bridges to link the molecules together. In ice, for example, each oxygen atom is surrounded tetrahedrally by four hydrogen atoms, as shown schematically in Fig. 1. Studies of the structure of ice by x-ray diffraction show that the oxygen atoms are separated from one another by a distance of 2.76 A, with a hydrogen atom in an intermediate position between them. Each oxygen atom is associated by covalent bonds wi hydrogen atoms about 1.00 A away, in a ith the empirical chemical formul

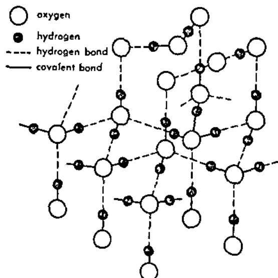


Fig. 1. Hydrogen bonding in structure of ice.

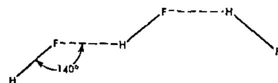


Fig 2. Hydrogen bonding in hydrogen fluoride.

H_2O. The remaining two hydrogen atoms are more remote (1.76 A), but these in turn are only 1.00 A distant from other oxygen atoms to which they are covalently bound. The hydrogen bond responsible for the association of water molecules in ice is the directed interatomic force which exists between a given oxygen atom and a hydrogen atom 1.76 A distant. The Pauli exclusion principle prohibits the formation of two covalent bonds to a hydrogen atom, and the hydrogen bond is therefore believed to be ionic in character.

Even in liquid water, hydrogen bonding persists, and there is some merit to the point of view that the entire liquid is one gigantic associated molecule. Individual H_2O units are known to move about independently in liquid water, however, so hydrogen bonding must be regarded as a dynamic equilibrium, in which bonds are continually being broken and reformed. Numerous bits of evidence for molecular association and hydrogen bonding exist. The high melting and high boiling points of water compared with those of hydrogen sulfide or hydrogen selenide, and its comparatively high latent heats of fusion and vaporization attest to its association. In liquid water, the infrared spectrum shows a broad absorption band near 3 μ, which is attributed to a characteristic vibration frequency of the hydrogen bond.

Hydrogen bonding is not limited to water, but is displayed by hydrogen fluoride in the liquid and vapor states, and by phenols, alcohols, and carboxylic acids, generally. In gaseous hydrogen fluoride, there is a tendency to form chains of associated molecules, and electron-diffraction studies indicate that the angle between the bonds is about 140° (Fig. 2). Moreover, in aqueous solutions of hydrogen fluoride, the HF_2^- ion is quite stable. Alcohols, phenols, and carboxylic acids in their liquid state give evidence of molecular association through hydrogen bonding from their x-ray diffraction patterns. Two diffuse rings are seen in these patterns, and the interatomic distance corresponds to one of them and varies according to the expected unit length in a chain. Figure 3 shows, in idealized form, the mode of hydrogen bridge formation in the acetic acid dimer.

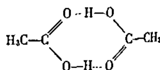
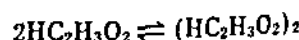


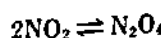
Fig. 3. Molecular association in acetic acid

In contrast to their behavior in aqueous solutions in which they are slightly dissociated into ions, organic acids are associated in solvents of low dielectric constant, such as benzene. Thus, acetic acid exhibits an anomalously high molecular weight in benzene, and approaches the value for the dimeric form in concentrated solutions, as required by the equilibrium



$$2HC_2H_3O_2 \rightleftharpoons (HC_2H_3O_2)_2$$

Di- and polymerization. Molecular association is not limited to molecules in which hydrogen bonding can occur. For example, nitrogen dioxide exists principally as a dimer in the vapor phase or in chloroform solutions at low temperatures. At low pressures or low concentrations and at elevated temperatures, however, it is largely dissociated into monomeric form:



$$2NO_2 \rightleftharpoons N_2O_4$$

Similar dissociation reactions are shown by aluminum chloride and ferric chloride, which exist as dimeric molecules (Al_2Cl_6 and Fe_2Cl_6) in the vapor phase and in organic solvents at low temperatures, but which are dissociated into monomers at elevated temperatures. Depending upon the point of view, these reactions may be regarded as either dissociation or association reactions. Similar considerations apply even to some elements, which may be partially or largely associated in the vapor state. Thus, sodium vapor contains a small concentration of Na_2 molecules, and arsenic vapor consists principally of As_4 molecules up to high temperatures. Even carbon vapor has been shown to contain the species C_2 and C_3 in addition to carbon atoms at high temperatures. See HYDROGEN BOND; LIQUID. [W.H.S.]

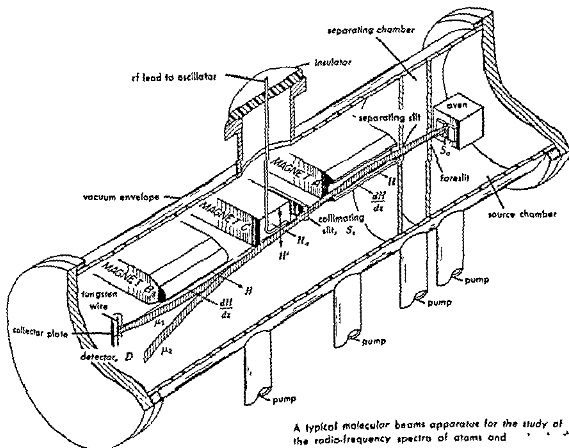
Molecular beams

The method of molecular or atomic beams is a powerful experimental technique for the investigation of a considerable range of physical phenomena

The method and the kinds of problems to which it is applied are essentially the same for atoms and molecules. The method is usually characterized by the production of an extremely well-defined beam of neutral atoms or molecules which is then subjected to further study by the application of one or more influences which modify the subsequent trajectory of the particles in the beam.

Techniques. The beam is most commonly generated by the evaporation of a salt or other substance from a cavity or oven whose internal volume is closed to the outside except for a narrow slit through which the material leaves (see illustration). The material of the oven is determined by the temperature at which evaporation is to occur and by the chemical properties of the substance to be evaporated. Beams of a large range of substances have been produced at a wide range of temperatures. For example, beams of mercury have been produced by evaporation from iron ovens at 400°K and of praseodymium by evaporation from a molybdenum oven, which holds the praseodymium in a thoria crucible, at 1840°K. In the case of molecular hydrogen it is desirable, in many studies, to cool the source to a convenient low temperature such as the boiling point of nitrogen (77°K). In important studies of atoms such as hydrogen, chlorine and bromine, which ordinarily occur only in molecular combination, the molecules are dissociated in a discharge tube, and the gas, which is then largely atomic, escapes through a slit

The particles escape from the source into a large solid angle, and a beam is defined by one or more additional slits, parallel to the source slit and in the path of the effusing material. In the absence of deflecting fields, the particles move in a straight line (the dashed line in the illustration). In typical setups, the width of the resulting beam ranges between 0.003 and 0.030 cm; the lengths, measured from source slit to detector, have ranged between 30 and 270 cm. In any case very strict requirements are imposed on the vacuum system, both in terms of the ultimate vacuum required to give a molecular mean free path (the average distance a molecule travels before colliding with another molecule) several times the length of the beam and in terms of the stability of the vacuum required to give an unfluctuating beam intensity at the detector. A common arrangement is shown in the illustration where the oven, a potential source of gas, is mounted in a separately pumped chamber which is in turn buffered from the main chamber through a separately pumped chamber. Typically the residual pressure in the main chamber should not exceed about 5×10^{-7} mm of mercury, though a lower pressure is required in some applications and a higher pressure is tolerable in others. For a specified beam width at the position of the detector and for a fixed rate of effusion from the oven, the beam



A typical molecular beams apparatus for the study of the radio-frequency spectra of atoms and

intensity falls off as the square of the distance from source to detector. Hence good design of an experiment involving molecular-beam techniques often calls for the achievement of a minimum over-all beam length.

All molecular-beam experiments require the measurement of the intensity of the beam and of changes in the intensity. Since the number of atoms or molecules of the beam incident on unit area in unit time at the detector is always much less than the number incident at the detector from the residual gas in the apparatus, it is necessary that a detector have a high preferential sensitivity to the particles in the beam. A number of ingenious schemes for giving preferential detection of a beam have been devised. The most common of these, applicable to a wide range of materials, depends on the fact that an atom with an ionization potential less than the work function of a metal will be evaporated from the metal at a suitable temperature as a positive ion. The constituents of the residual gas in a vacuum system are not ordinarily ionized upon impact with a heated tungsten wire. However, all of the alkali atoms, as well as gallium, indium, and thallium, will be ionized with an efficiency of almost 100% on a heated tungsten wire, which may be oxidized to increase the work function. Other atoms, like aluminum and barium, may be detected in the same way, but with less efficiency and more technical difficulty. A molecule that contains one of the atoms which may be detected in this way is also readily detectable. The detection of a beam of atoms or molecules that can produce ions consists then of the interposition in the path of the beam of a heated tungsten wire maintained at a small positive potential with respect to ground. The ions flow to a collector connected to ground through a high resistance and the current may be measured by conventional techniques.

The total number of particles intercepted by the detector is very small; a flux of 10^{8} particles/sec is fairly typical, though a much larger or smaller flux has been used in certain experiments. If every particle produces one ion, the current in a detector is about 1.6×10^{-11} amp. The time required to intercept 1 gram mole of the material at the detector is about 2×10^{8} years for a typical intensity.

Radio-frequency spectroscopy. The most fruitful application of the method of molecular beams has been as a spectroscopic device in the range of frequencies (about 10^{5}–10^{9} sec^{-1}) that can be generated by electronic means. The results of the first spectroscopic experiment by what is called the molecular-beam magnetic-resonance method were published in 1938 by I. Rabi and his colleagues; since then both experimental methods and experimental results have been increasingly elaborated and refined. For illustrative purposes only the initial, classic experiment, shown schematically in the illustration, will be discussed here.

Consider a diatomic molecule in which both the orbital and the spin angular momenta of the individual electrons add up in such a way that the net electronic angular momentum is equal to zero. The molecule then has no magnetic dipole moment arising from the electrons. Assume also that one of the two nuclei has a zero angular momentum and hence no magnetic moment but that the other nucleus has an angular momentum of $Ih/2\pi$ where h is Planck's constant and I is the spin quantum

is no other net source of magnetic dipole moment in the molecule, the nuclear dipole moment is the total magnetic dipole moment of the molecule. According to well-established principles of quantum mechanics, the nucleus can take on only two possible orientations with respect to a magnetic field, H. Either the spin is parallel to H, in which case the component of the spin along H is $m_I = ½$, or the spin is antiparallel to H, so that $m_I = -½$. The component of the magnetic moment of the nucleus and therefore of that of the molecule along the field is also positive in one case and negative in the other. The molecule exists, then, in two different energy states. These are characterized by the energies $W(m_I) = -(\mu_I/I)m_I H$, and since I is here taken to be ½, $W(m_I) = -2\mu_I m_I H$. A transition between the two energy levels involves a change of m_I from $+½$ to $-½$ or vice versa and the energy absorbed or emitted is $\Delta W = 2\mu_I H$. This energy is in the form of a photon of frequency f, $f = \Delta W/h = 2\mu_I H/h$. For the hydrogen nucleus f is about 40 Mc/sec at a field strength of $H = 10{,}000$ oersteds. Since the energy difference between the two levels is very small, even in the highest magnetic fields, spontaneous emission of a photon will not occur; but under the application of an oscillating magnetic field H', perpendicular to H, of frequency f and of suitable amplitude, the transitions from $m_I = +½$ to $-½$ and from $-½$ to $+½$ occur with equal probability. The measurement of f at which the transition occurs in a field H allows the determination of the nuclear magnetic moment μ_I.

Consider the system shown in the illustration. A beam of molecules issues from the slit S_o in the oven, is defined by the collimating slit S_c, and strikes the detector D. In the absence of any fields the line S_oS_cD is straight. If an inhomogeneous magnetic field, the A field, with a gradient dH/dz as shown is applied over a portion of the distance between S_o and S_c, molecules with $m_I = +½$ will be deflected along the direction of dH/dz and molecules with $m_I = -½$ in the opposite direction. Molecules that leave S_o along the line S_oS_c will not go through the slit S_c, but since molecules effuse from the oven at all angles to the line S_oS_c there will be some angle at which molecules leave S_o for which they will pass through S_c when deflected by the A field. This angle depends on the

molecular velocity. These molecules will not reach D in the absence of further fields. If between S_c and D there is placed an inhomogeneous magnetic field, the B field, for which H is in the same direction as that in A but in which dH/dz is in a direction opposite to that in A, all the molecules will be refocused on the detector for an appropriate choice of the various lengths along the trajectory. The refocusing is velocity independent; that is, it occurs for the complete range of thermal velocities that the molecules have as they leave S_o. In the illustration the trajectory of particles in a beam that has been deflected by the B field as described is shown by the ribbon marked μ_1. The particular trajectory shown corresponds to a positive magnetic moment. Other molecules, as they leave the source, will have a negative magnetic moment and other velocities but they will still be refocused on the detector. The pole tips on the A and B magnets (half of which are shown) are cylindrical surfaces and the gradient in B is reversed with respect to that in A by the simple expedient of inverting the pole faces.

The C magnet is placed between the A and B magnets. The field H_o of this magnet is uniform but in the same direction as that of the A and B magnets; its magnitude is arbitrary. Superimposed on H_o is an oscillating magnetic field H', perpendicular to H_o. If the frequency of the oscillating field is $f_o = 2\mu_I H_o/h$, transitions will occur from $m_I = +\frac{1}{2}$ to $m_I = -\frac{1}{2}$ and vice versa. The probability of transition is not the same for all molecules because of the spread in velocity, which exposes different molecules to the oscillating field for different lengths of time. In spite of this an average probability of transition of as much as 70% may be achieved. Since the sign of the magnetic moment has been reversed in the transition process, the direction of the force in the B field is reversed and molecules are deflected away from their original trajectories in such a way as to miss the detector. When a transition occurs the trajectory of the molecule in the first half of the system is unchanged, but since the sign of the magnetic moment is reversed by the second magnet, the direction of the force is reversed. The trajectory for a molecule that has made a transition is marked μ_2 in the illustration. If the intensity is measured as a function of the frequency of the oscillating field at fixed H_o, a minimum of intensity will occur at $f = f_o$. A knowledge of f_o and H_o then determines μ_I.

Spectroscopic applications. The system described is applicable not only to the simple molecule that has been assumed but to more complex molecules and especially to atoms. Atomic electrons do not ordinarily have a zero net angular momentum. The interactions of interest, then, are the interaction (1) of the total electronic magnetic moment (composed of contributions both from the spin and the orbital motion of the electrons) with an applied magnetic field, (2) of the nuclear magnetic moment with the applied field, and (3) of the nucleus with the electron configuration. In the case of molecules, the interaction of each of the nuclei with an applied magnetic field may be determined, as may the interactions of the nuclei with the electron configuration and with each other. Many molecules have a permanent electric dipole moment. The effective electric dipole moment depends on the state of the molecule, and the frequencies that characterize the transitions between two states may be found by a method analogous to that described if electric fields are substituted for magnetic fields.

The design of any spectroscopic experiment by the molecular-beam magnetic-resonance method capitalizes on a change in the component of the over-all dipole moment, magnetic or electric, along the direction of a field through which the molecule passes after it has made the transition to be observed. Very high accuracies in the measurement of line frequencies may be obtained; an accuracy of 1 part in 10^6 is commonplace. The frequency of the line between the two states of atomic cesium (Cs^{133}), for which the total electronic angular momentum is parallel and antiparallel to the spin angular momentum of the nucleus, has been given as $(9192.63183 \pm 0.00001) \times 10^6$ sec^{-1}; the uncertainty is about 1 part in 10^9.

The value of the molecular-beam magnetic-resonance method is that it can be used to measure intervals between energy levels that are very close together as compared to the relatively large intervals that separate the levels between which transitions giving rise to optical radiation occur. Closely spaced levels at zero magnetic field occur through the interaction of the spin-related properties of the nucleus with the electronic structure of the atom or molecule. At other than zero field, the splitting of energy levels is still importantly affected by nuclear properties. Thus the method has allowed the determination of a great many nuclear properties to be made with a very high order of accuracy. Among these properties are (1) the spin of the nucleus (the determination of the spins of relatively short-lived radioactive species has been one of the important recent lines of inquiry); (2) the magnetic dipole moment of the nucleus; (3) the electric quadrupole moment of the nucleus, that is, the departure of the nuclear charge distribution from sphericity; (4) higher order nuclear moments, such as the magnetic octupole moment. In addition, data on the size of the nucleus and on the distribution of magnetism in the nucleus have been obtained. The nature of the interactions between the nucleus and its environment has also been explored. For example, the anomalous magnetic moment of the electron has been determined by the molecular-beam magnetic-resonance method.

Spectroscopic experiments have been performed on the hydrogen atom in its metastable state and on the helium ion (He^+), also in its metastable state, for the purpose of investigating the details

of the electron-nucleus interaction in a two-particle system. In these experiments transitions are detected by capitalizing on the fact that certain substates of the metastable state may, under certain conditions, have a differential metastability with respect to others.

Nonspectroscopic applications. Prior to the first application of the method of molecular beams to spectroscopic studies, the technique had been employed for the study of a whole range of problems. O Stern and W. Gerlach used the technique in 1924 in a famous experiment that demonstrated the reality of space quantization. Considerable data on nuclear properties had been accumulated through a study of the deflections of atoms and molecules in inhomogeneous magnetic fields. Significant studies of the velocity distribution in gases, the scattering of molecules from surfaces and in gases, diffraction phenomena involving particles, and studies of other phenomena led the way to the development of the resonance method.

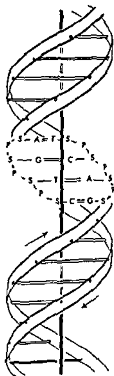
Recent studies of the velocity distribution of particles that effuse from an oven in which the alkali halides are evaporated have shown that at equilibrium in the gas phase, the typical alkali halide may contain large amounts of polymers of the simple diatomic molecule A study of the relative abundances of the various polymeric species as the temperature and pressure in the oven are varied leads to values of the heat of dissociation of the polymers. *See* MOLECULAR STRUCTURE AND SPECTRA; NUCLEAR MOMENTS; NUCLEAR STRUCTURE. [P.KU.]

Bibliography: P. Kusch and V. W. Hughes, Atomic and molecular beam spectroscopy, *Handbuch der Physik*, vol. 37, 1958; N. F. Ramsey, *Molecular Beams*, 1956

Molecular biophysics

A branch of science which analyzes various biological phenomena on their most fundamental level, that is, on the level of the behavior of molecules. Because the primary emphasis is on a molecular explanation or description, one of the major tools in this discipline is chemistry, which is the science of molecules On the other hand, many of the techniques used in molecular biophysics are taken directly from physics, such as the use of x-ray diffraction or nuclear magnetic resonance to study molecular structure. Thus molecular biophysics can be characterized as a discipline which utilizes the selective aspect of both chemistry and physics to study phenomena of biology on the molecular level.

The borderlines of this discipline merge almost imperceptibly into various other branches of science. Thus there is a large area of common ground with what has traditionally constituted biochemistry. For example, some subjects such as the synthesis of large molecules and the mechanism of enzyme action are certainly studied in both biochemistry and molecular biophysics. However, an area such as nutrition has very little representation



Schematic diagram of structure of the deoxyribonucleic acid molecule, a polymer coiled into a two-stranded helix with sugar deoxyribose (S) and a phosphate group (P) joined to form the backbone These are held together by the purines, adenine (A) and guanine (G), which are hydrogen-bonded to the pyrimidines, thymine (T) and cytosine (C). Axis of the molecule is also shown by a vertical bar.

in the field of molecular biophysics. Molecular biophysics is distinguished from biochemistry by the kind of question the investigator is trying to answer. For example, a large part of biochemistry traditionally has been directed towards the elucidation of various metabolic routes and cycles which take place in the living cell. In these, a series of reactions is sequentially performed upon a given molecule, resulting in a conversion of chemical energy. For example, the glucose molecule is degraded into smaller molecular units so that its chemical energy can be transformed into other molecules which can be utilized by the cell in biosynthetic and other functions. The area of classical biochemistry has traditionally been concerned with the identification of the substances involved in these metabolic cycles and the reactions in which they are engaged. In molecular biophysics the scientist would be more apt to ask why these reactions occur, and would look for an answer in terms of the thermodynamics and kinetics of the chemical reaction or in terms of the molecular structure of the products and reactants. In the field a great deal of effort is spent on understanding the molecular structure and reaction mechanism at the active site at which these metabolic reactions occur. Thus a great deal of activity in molecular

biophysics is common to the field of biochemistry; however, the approach is somewhat different and the kind of question answered is also different.

One of the best ways of characterizing molecular biophysics is to outline briefly the subject matter of the field and discuss some of the active areas of research. For this purpose, four general divisions of the subject will be considered.

Molecular structure and function. At the core of this area of molecular biophysics is the problem of the structure of molecules which are found in biological systems. In this context, the term structure means the distribution of the atoms in space, that is, a three-dimensional picture of the location of the atoms in the molecule. The major tool for this type of investigation up to the present time has been x-ray diffraction. Since x-rays have a wavelength which is close to the distances between atoms in molecules, it is possible to interpret the interference produced by the scattering of this radiation to determine molecular structures. X-rays are scattered by electrons and accordingly from these studies a three-dimensional electron density plot can be obtained. An electron density plot has on it the position in space of all the atoms in the molecule and this fact has a substantial bearing on the problem of the chemical reactivity of the molecule since the energetics and the reactivity are direct consequences of the organization of the constituent atoms. Another important feature is the fact that structure and function are closely related to each other on the molecular level. An example of this may be seen in the exciting work which has been undertaken on the determination of the structure of proteins utilizing x-ray diffraction techniques. The results of these investigations are a complete three-dimensional description of protein molecules, which are among the most complicated molecules in biological systems. It is reasonable to believe that this three-dimensional description of the proteins will lead rather directly to an understanding of the mechanism of enzyme action. Thus the mechanism by which a small substrate molecule interacts with the larger surface of the protein enzyme will be clarified. Because of the distribution of charged groups, hydrogen-bonding atoms, and hydrophobic side chains, the nature of the catalytic surface and the mechanism for inducing the chemical reaction will be elicited. In this way the function of the biological catalyst will be understood almost as a direct consequence of understanding its structure. *See* ENZYME; PROTEIN; X-RAY DIFFRACTION.

In this area various other experimental techniques are of great value. Infrared spectroscopy is of importance because it gives information about the vibrational motions of the molecule. Similarly, ultraviolet spectroscopy informs the investigator about the various electronic energy levels which the molecule has and these are of importance in understanding how energy can be transported from one molecule to another.

Electron microscopy has in recent years become a very powerful tool in the field of molecular biophysics because it enables the investigator to visualize the form of individual molecules. At the present time the individual molecules which can thus be studied are very large since the resolving power of most microscopes now in use is not much better than 10 A. However, there is reason to believe that this limit of resolution will be improved. By using this tool various details of the structure of viruses are now being seen. A recent example of this is the discovery that the outer shell of most spherical viruses consists of an aggregation of smaller protein molecules which are now directly resolvable in the electron microscope. Similarly, the electron microscope has proven very useful in showing the size and shape of very elongated molecules such as the collagen molecule which is found in connective tissue, or the macromolecular nucleic acids. *See* CONNECTIVE TISSUE; MICROSCOPE, ELECTRON; NUCLEIC ACID.

Many other techniques are valuable in molecular biophysics Among the more important are the various methods for studying macromolecular properties. These include light scattering, ultracentrifugation, double refraction of flow, optical rotatory studies, viscometry, and osmometry In general, any technique of physical chemistry can be applied to the study of biological systems on the molecular level. *See* OPTICAL ACTIVITY; PHYSICAL CHEMISTRY; ULTRACENTRIFUGE.

Energy transport and conversion. A fundamental property of a living cell is its ability to absorb energy from its environment and alter the form of the energy in various ways so that it can be utilized to carry out a variety of chemical reactions which are necessary for the maintenance of life. An important area of molecular biophysics is the study of the molecular mechanisms in photosynthesis. Recent work in this field, utilizing the technique of electron spin resonance has shown that free radicals are generated in photosynthetic particles when they are active. It appears that the organized photosynthetic particles function in a way which can be understood only because of recent advances in the field of solid-state physics. *See* PHOTOSYNTHESIS; SOLID-STATE PHYSICS.

Most of the research work done at the present time in the field of energy transport and conversion has not yet been amenable to study on the molecular level. Although some work has been done on photoreception in the eye, very little is known about the molecular mechanism underlying energy transport between molecules. There are some suggestions that exciton transfer is an important part of energy transport in organized aggregates, but most of this work is still in a very early stage of development.

The conversion of chemical energy into mechanical energy is carried out by muscles. The electron microscope has been of great use in describing the manner in which various large molecular rods slide

over each other during contraction, but the basic reaction underlying this movement is still the object of intensive research. See MUSCLE; MUSCLE (BIOPHYSICS).

Many studies have been made of the thermodynamics of the cell and its components. It is apparent that a living cell is an entropy-producing machine, in that it takes in molecules which are highly organized and, in general, breaks them down into smaller components. However, various parts of the cell carry out reactions which result in a localized decrease of entropy as may be seen in the biosynthetic reactions leading to the production of macromolecular proteins or nucleic acids. See ENTROPY.

Conversion and transport of matter. Living organisms have a continual flux of material substances passing through them which are modified by the biological system to maintain its integrity. The molecular approach to the problems of the conversion of one molecule into another has had a relatively short history. As mentioned previously, a point is being approached at which an understanding of the mechanisms of these conversions through the important action of the catalytic enzymes will be possible. This represents a very active area of molecular biophysics.

On a much more complicated level, a consideration can be made of the problems of osmotic regulation or the more generalized membrane problem. All organisms are divided into compartments by membrane structures which have very specific properties of allowing certain ions or molecules to pass through in one direction but not in another. Such membranes are found on all cells, and have very special properties in nerves. However, subcellular membranes also exist, such as those which cover mitochondria or the nucleus. Broadly speaking, the molecular basis for their specificity is largely unknown at the present time. A considerable amount of study is being done on membrane behavior and molecular structure, especially in nerve cells.

Information transfer. A biological system is characterized not only by the complexity of the chemical reactions which it carries out but also by the fact that these reactions are highly coordinated. This condition requires information transfer on many levels.

On the level of complicated systems, molecular knowledge is scant, since very little is known regarding the mechanism which occurs at the synapses in a nerve net and which determines whether or not an impulse passes through the synapse. In addition, virtually nothing is known about the molecular mechanisms which take place in the central nervous system on the more highly organized levels of nervous activity. Indeed it might be said that this field will represent one of the most promising areas of future expansion of the field of molecular biophysics. See BIOPOTENTIALS AND ELECTROPHYSIOLOGY.

On a somewhat simpler level, in the field of genetics, the investigator is starting to learn a great deal about information transfer on the molecular level. There is now a reasonably good understanding of the molecular mechanism which is used for transmitting from mother cell to daughter cell the genetic information which governs the development of the cell, and in turn determines the type of protein molecules which are synthesized in a biological system. It is worthwhile discussing this example of information transfer in a little more detail since it will bring in many areas of molecular biophysics. In the illustration the deoxyribonucleic acid molecule has the form of a two-stranded helix in which the backbone chains of sugar phosphate groups run in opposite directions as indicated by the arrows in the diagram. The two helically wound strands of sugar-phosphate residues are held together by a series of purine pyrimidine hydrogen-bonded pairs which have the unusual property of specificity. Thus the purine adenine hydrogen bonds uniquely with the pyrimidine thymine, while the purine guanine hydrogen bonds with the pyrimidine cytosine. The molecular structure of this two-stranded helical molecule has been worked out in great detail so that the distribution in space of all the component atoms is known.

The molecule of deoxyribonucleic acid is the major carrier of genetic information in biological systems. Thus it is one of the major constituents on chromosomes and it comprises the information-carrying element in the nucleus of cells; as such it is the molecule which directs both the development of the system and indirectly directs the synthesis of the various types of molecules found in the organism. In the field of molecular genetics a great deal of work has been done to define the relationship between the nucleic acid molecule and the functional units of activity, that is, the genes or the subunits of genes which are responsible for various inheritable characteristics. In recent studies using the bacteriophage virus as a test system, it has been possible to carry this genetic analysis down to the level which indicates that adjoining functional units on a DNA molecule must be separated by only a small number of purine pyrimidine base pairs on the DNA molecule. See DEOXYRIBONUCLEIC ACID; GENETICS.

One of the problems of central concern is the question of how the genetic information is replicated or, on the molecular level, how the DNA molecule replicates. The clue to understanding this question has come from a study of the molecular structure of DNA, shown in the figure. Thus the two-stranded molecule may be described as being made of complementary halves, in that one half of the molecule defines in a specific way the complementary opposite half of the molecule. Recent work on the mechanism for replicating deoxyribonucleic acid indicates that the replication of the DNA molecule occurs by utilizing one half of the molecule, a single strand, which acts as a template for organizing on itself its complementary mate. Thus in the over-all process there is an initial separation of the two helical strands of the de-

oxyribose molecule followed by a synthesis of the complementary halves of the molecule. In this way there has been a complete system of molecular replication. This is an example of information transfer on the molecular level in that the initial system of one molecule has replicated to produce two identical molecules which contain the same information. In this case, the information is carried in the particular sequence of purine and pyrimidines along an individual helical strand. It is quite likely that there is another molecular mechanism, unknown at present, whereby this information is transmitted, probably to another nucleic acid molecule in the cell and ultimately to the specific protein molecules which are synthesized by this genetic material. This very active area of molecular biophysics deals with the problem of information transfer, with problems of molecular structure and function, and with the problem of the transport and conversion of matter. In this respect it is a good example of the unification which arises in a fundamental study in the field of molecular biophysics [A.B.]

See BIOPHYSICS.

Molecular electronics

A concept of energy transformation in which molecular and atomic crystals are used as the building blocks of solid circuits. Atomic spins, fields, and charges are employed to perform the fundamental functions normally accomplished with circuits using standard electronic components.

Approach. The progress in molecular electronics is due to the cooperative efforts of several branches of science. Electronic engineers with a broad background in molecular engineering specify the requirements of the system. Mathematicians and theoretical physicists develop a system transfer equation in terms of molecular logic functions and devise the various material domains required to achieve these functions. Solid-state physicists and molecular engineers then create the molecular electronic functional unit.

This joint effort emphasizes the new approach required for molecular electronic-circuit development. Rather than by use of the techniques of conventional circuitry, the desired functions are accomplished by control of energy flow through arrangement of the atomic distribution of material. For example, conventional circuits use an array of inductors and capacitors for tuning; with molecular methods, the transit-time effect of the charge carriers and the energy storage capability of the material can be used instead.

Advantages. The techniques of conventional miniaturization have been successful in reducing the volume and weight of conventional electronic systems, but further progress is limited as long as such usual circuit components as resistors, inductors, and capacitors are used. Molecular electronic circuits permit further substantial reductions in both volume and weight. As an example, a light telemetry subsystem has been developed to sense light radiation and produce a high-frequency signal proportional to radiation intensity. A conventional transistorized electronic circuit to perform the same function weighs 7 grams (g) and occupies a volume of 1 cubic inch (in.^{3}) The molecular electronic unit weighs 0.02 g and has a volume of 0.001 in.^{3}

Another potential advantage of molecular electronic devices is their greater reliability when compared to conventional circuits. This advantage comes largely from a reduction in the number of individual components and their associated soldered joints. In the above example, the conventional circuit contained 15 soldered connections, while the molecular electronic device contained only 2 soldered connections.

A further advantage is that molecular electronic units can be designed to operate at substantially higher efficiencies than either vacuum-tube or transistor circuits. If an output power of 50 milliwatts (mw) is desired from the circuit in the previous example, the transistorized version requires 750 mw of input power, while the molecular electronic device requires only 60 mw.

At the present time, molecular electronic circuits also have certain disadvantages. One disadvantage is the present high cost. Most of the molecular electronic devices that have been developed to date have been built in the laboratory, and the investment in research time has been large. Some manufacturers feel that costs may remain high even when commercial production is started, until the yield of acceptable units is increased.

Another disadvantage, shared by most solid-state electronic devices, is their susceptibility to damage from nuclear radiation Where the molecular electronic unit is to be subjected to high radiation levels, adequate shielding must be provided. In some cases, the volume and weight of this shielding nullify one of the important advantages of molecular electronic circuits

Applications. Present development work is aimed at producing many different types of functional circuit units. These units can then be combined in various ways to duplicate the functions of more complicated electronic circuits. Some of the functional circuit units which have already been developed are (1) a 5-watt audio amplifier with a frequency range from zero to 20 kc, measuring ¾ in. in diameter; (2) a 40-watt direct-current amplifier whose output is controlled by about 4 ma of input current; (3) a tuned amplifier, using a semiconductor notch filter, whose frequency can be varied by changing the applied voltage; (4) a multistage flip-flop circuit, which can operate at frequencies as high as 1 Mc and with a current gain per stage greater than unity; and (5) a variable potentiometer with no moving parts, which generates a signal proportional to the product of two signal inputs.

These examples represent only the first preliminary results from present research in molecular electronics. As further basic units are developed, functions of more complicated conventional

electronic circuits can be duplicated. *See* MICROCIRCUITRY; MINIATURIZATION OF EQUIPMENT; SOLID-STATE PHYSICS. [P.J.K.]

Molecular physics

Modern molecular physics includes the quantum mechanical explanation of the several kinds of chemical binding between the atoms in a molecule, directed valence, the polarizability of molecules, the quantization of the molecule's vibrational, rotational, and electronic motions and the phenomena arising from intermolecular forces. The study of molecular spectra has contributed greatly to man's knowledge of molecular structure and dynamics. In turn, the numerous agreements between the detailed features of these spectra and the predictions of quantum mechanics constituted some of the earliest proofs of the correctness of this formulation of quantum theory. Microwave spectroscopy and molecular beams experiments have produced measurements of high accuracy on the fine structure of certain of these spectra, resulting in the determination of nuclear moments and spins, as well as of atomic masses.

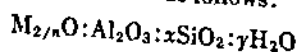
Intermolecular forces are involved in the liquefaction of gases, adhesion, cohesion, surface tension, and transport phenomena such as diffusion, viscosity, and thermal diffusion. Since about 1930 there has been a marked advance in the theoretical interpretation of all of these related matters, using quantum mechanical methods and knowledge of the structure of ...

domain of molecular physics. For a representative sampling, *see* CHEMICAL BINDING; INFRARED SPECTROSCOPY; INTERMOLECULAR FORCES; MICROWAVE SPECTROSCOPY; MOLECULAR BEAMS; MOLECULAR STRUCTURE AND SPECTRA; VALENCE. [W.W.W.]

Molecular sieve

One of a series of unique adsorbents which are crystalline metal aluminosilicates belonging to a class of minerals known as zeolites. These minerals are found widely scattered in nature in relatively small quantities. Synthetic forms of the naturally occurring minerals, as well as many species having no known natural counterpart, have been prepared by a hydrothermal process. An important characteristic of the zeolites is their ability to undergo dehydration with little or no change in crystal structure. The dehydrated crystals are honeycombed with regularly spaced cavities interlaced by channels of molecular dimensions which offer a very high surface area for the adsorption of foreign molecules.

Structure. The basic formula for all crystalline zeolites can be represented as follows:



$$M_{2/n}O:Al_2O_3:xSiO_2:yH_2O$$

where M represents a metal ion and n its valence. In general, a particular crystalline zeolite will have values for x and y that fall in a definite range. For example, two of the synthetic varieties of molecular sieves which are used commercially are designated as type A and type X. For type A, the value of x is about 2.0; for type X, the value of x is about 2.5. When fully dehydrated, the value of y for both types is zero. The crystal structure of molecular sieves consists basically of a 3-dimensional framework of SiO_4 and AlO_4 tetrahedrons (Fig. 1). The tetrahedrons are cross-linked by the sharing of oxygen atoms, so that the ratio of oxygen atoms to the total of silicon and aluminum atoms is equal to two. The electrovalence of the tetrahedrons containing aluminum is balanced by the inclusion of cations in the crystal. One cation may be exchanged for another by the usual ion-exchange techniques. The size of the cation and its position in the lattice determine the effective diameter of the pore in a given crystal species. The influence of various cations on the effective pore size of the molecular sieve type A is shown by the effective pore diameters of the potassium ion, K^+, sodium ion, Na^+, and calcium ion, Ca^{++}, which are approximately 3, 4, and 5 angstroms (A), respectively.

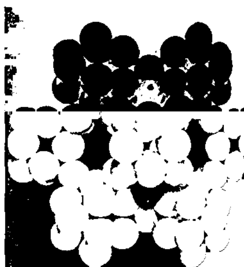


Fig. 1. Molecular sieve type A crystal model. Dark spheres represent the included cations, and light spheres the SiO_4 or AlO_4 tetrahedrons.

The crystal habit of molecular sieve type X is similar to that of diamond in which the carbon atoms are replaced by silica-alumina polyhedrons. With alkali metal ions present in the structure the effective pore diameter is 9–11 A; with the alkaline-earth cations present, the effective diameter is 8–9 A.

Properties. The properties of molecular sieve adsorbents which distinguish them from nonzeolitic adsorbents are the relatively strong forces existing at the adsorption surface and the uniform pore size which is controlled, in a given crystal species, by the associated cation. The strong surface forces

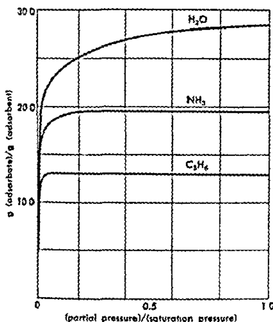


Fig. 2. Isotherms for type 5A molecular sieves at 25°C.

are reflected in the peculiar shape of the adsorption isotherm, the character of the adsorption isobar, and the relatively high heats of adsorption. The isotherm, in which the capacity for a given adsorbate is plotted against pressure or concentration at constant temperature, is of the so-called Langmuir type (Fig. 2). The shape of the isotherm is approximately rectangular, rising steeply at low partial pressures or concentrations and leveling off when maximum loading is attained. The isobar, in which capacity is plotted against temperature at constant pressure, shows that molecular sieves have an unusually high capacity at elevated temperatures. The relatively high exothermic heats which are characteristic of adsorption on molecular sieves necessitate somewhat higher heat requirements to effect desorption than are required with other adsorbents.

The characteristics of molecular sieves which have been described can be utilized in a variety of applications. Their unique adsorption properties make them particularly useful for the drying, purification, and separation of gases and liquids. Used conversely, the molecular sieve can be preloaded with chemical agents which remain isolated from a reactive system in which they are dispersed, until released from the adsorbent either thermally or by displacement with a more strongly adsorbed compound. The presence in the crystal lattice of an associated, exchangeable metal ion provides the basis for use as a cation-exchange medium. Their chemical composition and crystal structure makes them novel catalysts and catalyst supports.

As predictable from the water-adsorption isotherm (Fig. 2), molecular sieves are capable of drying gases and liquids to extremely low residual water concentrations. The isotherm also shows that, even at low initial water concentrations in a gas or liquid, the capacity of the desiccant is relatively high.

By virtue of the uniform pore size of a given molecular sieve crystal type, molecules having a minimum projected cross section larger than the effective diameter of the zeolite pore are excluded from the internal surface. Molecules having a minimum projected cross section smaller than the effective pore diameter are adsorbed internally. This phenomenon is utilized in separating molecules of fluid mixtures on the basis of their size or shape. For example, molecular sieve type 5A has an effective pore size such that straight-chain hydrocarbons, which can enter the pores, are adsorbed. Branched-chain and cyclic hydrocarbons, on the other hand, are excluded from the pore system and thus are effectively separated from the straight-chain compounds.

In general, when two molecules of similar volatility are small enough to enter the pore system, separation is based on the degree of unsaturation or the polarity of the molecules. The more unsaturated or the more polar molecule is more strongly adsorbed. See ADSORPTION; ION EXCHANGE; ZEOLITE. [F.M O.; R.L.MA.]

Bibliography: D. W. Breck et al., Crystalline zeolites: I, The properties of a new synthetic zeolite, type A, *J. Am. Chem. Soc.*, 78(23):5963–5971, 1956; E. S. Dana, et al., *System of Mineralogy*, 2 vols., 7th ed., 1944, 1951; T. B. Reed and D. W. Breck, Crystalline zeolites: II, Crystal structure of synthetic zeolite, type A, *J. Am. Chem. Soc.*, 78 (23):5972–5977, 1956.

Molecular structure and spectra

Until the advent of quantum theory, ideas about the structure of molecules evolved gradually from analysis and interpretation of the facts of chemistry. Chemists developed the concept of molecules as built from atoms in definite proportions, and identified and constructed (synthesized) a great variety of molecules. Later, when the structure of atoms as built from nuclei and electrons began to be understood with the help of quantum theory, a beginning was made in seeing why atoms can combine in definite ways to form molecules; also, infrared spectra began to be used to obtain information about the dimensions and the nuclear motions (vibrations) in molecules. However, a fundamental understanding of chemical binding and molecular structure became possible only by application of the present form of quantum theory, called quantum mechanics. This theory makes it possible to obtain from the spectra of molecules a great deal of information about the nature of molecules in their normal as well as excited states, and about dissociation energies and other characteristics of molecules. For an important aspect of molecular structure which is treated separately, see CHEMICAL BINDING.

Molecular sizes. The size of a molecule varies approximately in proportion to the numbers and sizes of the atoms in the molecule. Simplest are diatomic molecules. These may be thought of as built of two spherical atoms of radii r and r', flattened where they are joined. The equilibrium value R_e of the distance R between their nuclei is then smaller than the sum of the atomic radii (Fig. 1). However, the nuclei of atoms in two different molecules cannot normally approach more closely than a distance $r + r'$; r and r' are called the van der Waals radii of the atoms. The smallest molecule is hydrogen (H_2), with two electrons whose negative charges just equal the positive charges of the two nuclei. Here r is about 1.2 angstroms (1 A = 10^{-8} cm) giving $r + r' = 2.4$ A, but R_e is only 0.74 A. In HCl (hydrogen chloride), $r = 1.2$ for H and 1.8 A for Cl, but R_e is only 1.27 A.

To describe a polyatomic molecule, one must specify not merely its size but also its shape or configuration. For example, carbon dioxide (CO_2) is a linear symmetrical molecule, the O—C—O angle being 180°. The H—O—H angle in the nonlinear water (H_2O) molecule is 105°. Many molecules which are essential for life contain thousands or even millions of atoms. Proteins are often coiled or twisted and cross-linked in curious ways which are important for their biological functioning.

Dipole moments. Most molecules have an electric dipole moment. In atoms, the electron cloud surrounds the nucleus so symmetrically that its electrical center coincides with the nucleus, giving zero dipole moment; in a molecule, however, these coincidences are disturbed, and a dipole moment usually results.

Thus, when the atoms of HCl come together, there is some shifting of the H-atom electron toward the Cl. A complete shift would give H^+Cl^-, which would constitute an electric dipole of magnitude eR_e, where e is the electronic charge. But in fact the dipole moment is only $0.17\ eR_e$. This is because the actual electronic shift is only fractional. For further discussion, *see* ELECTRONEGATIVITY.

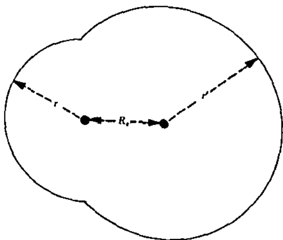


Fig 1. Diatomic molecule with nuclei at distance R_e apart, built from atoms of *radii* r and r'.

Although in molecules such as H_2, N_2, and CO_2 partial shifts of electronic charge from the original atoms do take place, these necessarily occur so symmetrically that no dipole moment results. Many larger molecules also have zero dipole moments by virtue of high symmetry. Examples are methane (CH_4), uranium hexafluoride (UF_6), and benzene (C_6H_6).

In general, the dipole moment of a neutral molecule is defined as the vector sum of quantities $+QS$ for the nuclei and $-es$ for the electrons. Here, Q is the charge on any nucleus and S its vector distance from any fixed point in the molecule, while s is the average vector distance of any electron from the same point. To calculate a dipole moment using these definitions, quantum mechanics must be used.

However, a study of what is known experimentally about molecular dipole moments has led to useful semiempirical generalizations. Bond moments and group moments have been obtained for various types of chemical bonds and of chemical groups or radicals. By adding these vectorially, the actual dipole moment of a large molecule can often be reproduced fairly accurately. In CH_4 or CO_2 one can assume a moment for each C—H or C═O bond, even though these cancel out vectorially to give a zero resultant. In the linear molecule OCS, the unequal moments of the C═O and C═S bonds give a nonzero resultant. Because of the zero moment of CH_4, the CH bond moment and the CH_3 group moment must be equal and opposite. In CH_3Cl, the total moment can be thought of as the vector sum of the H_3C group moment and the C—Cl bond moment. For additional information, *see* DIELECTRIC CONSTANT.

When molecules vibrate, their dipole moments usually vary. Figure 2 shows how the dipole moment μ in a diatomic molecule may vary with R; the quantity previously discussed is μ_e, the value of μ at R_e. When μ_e is zero because of symmetry, it remains zero for symmetrical vibrations but, in polyatomic molecules, varies during unsymmetrical vibrations.

Molecules may possess magnetic as well as electric dipole moments; for information on the former, *see* MICROWAVE SPECTROSCOPY.

Molecular polarizability. In the preceding consideration of dipole moments, the discussion has been in terms of atoms and molecules free from external forces. An electric field pulls the electrons of an atom or molecule toward it and pushes the nuclei away, or vice versa. This action creates a small induced dipole moment, whose magnitude per unit ... polarizability.

[illegible] ...pressed as ... corrections ... t. Polariza... series as F.

Cl, Br, I, and also from HF to HI, or F_2 to I_2.

Molecular polarizabilities can also be expressed approximately as sums of bond polarizabilities. These are anisotropic, being greater along bonds than perpendicular to bonds.

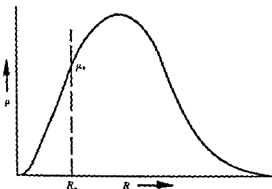


Fig 2. Electric dipole moment μ of typical diatomic molecule as function of internuclear distance R; μ_e is the dipole moment at the equilibrium distance R_e

Molecular energy levels. The states of motion of nuclei and electrons in a molecule, or of electrons in an atom, are restricted by quantum mechanics to special forms with definite energies. The state of lowest energy is called the ground state, all others are excited states. In analogy to water levels, one speaks of energy levels. Excited states exist only momentarily, following an electrical or other stimulus. *See* QUANTUM CHEMISTRY; QUANTUM MECHANICS.

Energy levels are either discrete or continuous. The levels of a self-contained atom or molecule are restricted to special, sharply defined values (discrete levels). When an atom or molecule is ionized, that is, when one of its electrons has enough energy to escape completely, the energy can take on any value exceeding the minimum escape energy. One then speaks of continuous levels or of an ionization continuum. Molecules also have dissociation continua, which are discussed later in this article.

Excitation of an atom consists in a change in the state of motion of its electrons. Electronic excitation of molecules can also occur, but alternatively or additionally, molecules can be excited to discrete states of vibration and rotation.

In a diatomic vibration, R varies periodically above and below R_e. The possible vibration energies E_v are given by

$$E_v = hc\omega_e[(v + \tfrac{1}{2}) - x_e(v + \tfrac{1}{2})^2] + \cdots \quad (1)$$

Here $c\omega_e$ is called the small-amplitude vibration frequency, and h is Planck's constant (6.62×10^{-27} erg-sec); x_e is a small quantity which is nearly always positive. The vibrational quantum number v can take whole-number values 0, 1, 2, etc. The $+ \cdots$ in Eq. (1) indicates small correction terms. The *zero-point vibration energy* $\tfrac{1}{2}hc\omega_e(1 - \tfrac{1}{2}x_e)$ present in the ground vibration state ($v = 0$) is a characteristic manifestation of quantum theory.

The value of $c\omega_e$ depends on the masses m_1 and m_2 of the atoms and the force constant k

$$c\omega_e = \sqrt{k[(1/m_1) + (1/m_2)]} \quad (1a)$$

The frequency $c\omega_e$ (c = speed of light) is so written for reasons of convenience in spectroscopic work, where the factor c is usually dropped.

The quantities R_e, k, and the dissociation energy D are the most important properties of a potential curve, which shows how the energy of attraction $U(R)$ of the atoms varies with R; k is d^2U/dR^2 taken at R_e. The $U(R)$ curve and vibrational levels for the ground electronic state of H_2 are shown in Fig 3. Similar curves, but with other R_e, k, and D values, exist for other electronic states and other molecules. Molecules have also repulsive electronic states, whose $U(R)$ curves rise steadily with decreasing R. These are often important for spectroscopy and in atomic collisions. For stable (attractive) $U(R)$ curves, the vibrational levels decrease in spacing as v increases, until finally, as the spacing approaches zero, a maximum v is reached; in Fig 3 this is 14. After a small gap, a dissociation continuum of energy levels then sets in. Here the atoms have enough mutual kinetic energy to fly apart. For repulsive states, there is only a dissociation continuum, with no vibrational levels. Figure 4 illustrates how strongly vibration level spacings can vary: both k and $1/m$, and so $c\omega_e$, decrease from H_2 to O_2 to I_2. Figure 4 also illustrates the effect of mass in isotopic molecules.

The total energy of any molecule can be written

$$E = E_{el} + E_v + (E_r + E_{fs} + E_{hfs} + E_{ext}) \quad (2)$$

Both the electronic energy E_{el} and vibration energy E_v can be discrete or continuous. The quantities E_r, E_{fs}, and E_{hfs} denote rotational, fine structure, and hyperfine structure energies. The latter two appear as small or minute splittings of the rotation levels. The spacings ΔE of adjacent discrete levels of each type are usually in the order

$$\Delta E_{el} \gg \Delta E_v \gg \Delta E_r \gg \Delta E_{fs} \gg \Delta E_{hfs}$$

The fine structures of rotational levels differ

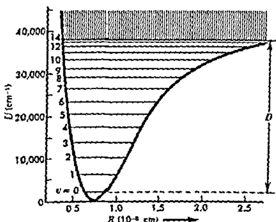


Fig 3. $U(R)$ curve of ground electronic state of H_2 with vibrational levels and dissociation continuum. D indicates the dissociation energy. (After G. Herzberg, Molecular Spectra and Molecular Structure, vol. 1, ed., Van Nostrand, 1950)

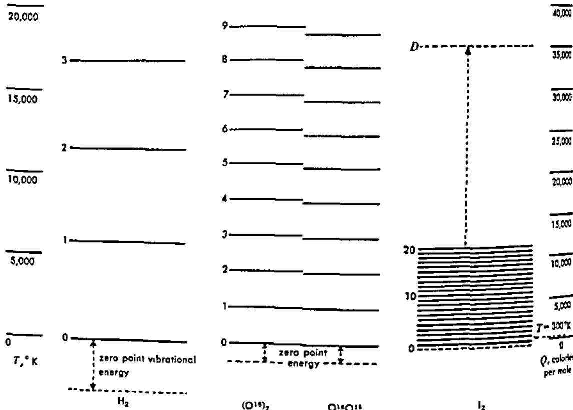


Fig. 4. Lowest vibrational levels of H_2, O_2, and I_2, numbered by vibrational quantum number v. Level spacings decrease with increasing v. Where spacing reaches zero, the molecule dissociates, dissociation level D is indicated for I_2. Energies are given by scale at right. Scale at left shows average energy of vibration at various temperatures.

strongly for different types of electronic states. The simplest diatomic electronic states are called $^1\Sigma$ states, and include $^1\Sigma^+$ and $^1\Sigma^-$ types for heteropolar, and $^1\Sigma^+_g$, $^1\Sigma^+_u$, $^1\Sigma^-_g$, and $^1\Sigma^-_u$ for homopolar molecules. Most even-electron diatomic and linear polyatomic molecule ground states are $^1\Sigma^+$ states ($^1\Sigma^+_g$ if homopolar). The rotational levels of $^1\Sigma$ states have no fine structure; hyperfine structure, because of interaction with nuclear spins, is usually on too small a scale to detect by optical ...

... LINES); HYPERFINE STRUCTURE; STARK EFFECT; ZEEMAN EFFECT.

The rotational levels of any $^1\Sigma$ state are given by

$$E_r = hcB_vJ(J+1) + \cdots \quad (3)$$

The quantity B_v is related to the moment of inertia I [$I = m_1m_2R^2/(m_1+m_2)$], and to v, by

$$B_v = (h/8\pi^2c)\,\overline{(1/I)}_v = B_e - \alpha_e(v + \tfrac{1}{2}) + \cdots = B_0 - \alpha_e v + \cdots \quad (3a)$$

The rotational quantum number J can have any whole number value from 0 up, and corresponds to an angular momentum $(h/2\pi)\sqrt{J(J+1)}$. The averaging of $1/I$ in Eq. (3a) normally results in a slow decrease of B with increasing v (α_e is usually a small positive quantity). The quantity B_e refers to a hypothetical nonvibrating molecule ($R = R_e$).

Figure 5 illustrates how enormously rotational level spacings can vary because of differences in m and R_e (both are much greater for I_2 than H_2). The effect of mass for isotopic molecules is illustrated for O_2. Comparison with Fig. 4 illustrates the relation $\Delta E_v \gg \Delta E_r$ mentioned earlier.

Polyatomic molecules have much more complicated patterns of vibrational and (usually) of rotational energy levels than diatomic molecules. The number of normal modes (independent forms) of vibration for a molecule with n atoms is $3n - 6$ for nonlinear, and $3n - 5$ for linear molecules. Each normal mode is a cooperative vibration of some or all the atoms moving with the same frequency, characteristic of the mode. Sometimes two or even three modes are so related in form that their frequencies are identical. These are called degenerate vibrations.

Figure 6 depicts the normal modes of H_2O and CO_2. They are labeled by symbols which also denote their frequencies. The arrows indicate the directions of motion of the atoms during one phase of vibration. The CO_2 frequency ν_2 is twofold degenerate: there are two independent modes corresponding to motion in either of two planes at right an-

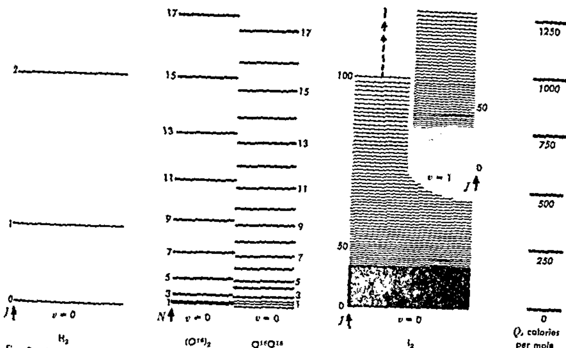


Fig. 5. Lowest rotational levels of H_2, O_2, and I_2. For H_2 and I_2, J is the rotational quantum number, Eq. (3). O_2 is in a Hund's case *b* triplet state, and the rotational levels are designated by N, where the total angular momentum $J = N + 1$, N, and $N - 1$. This narrow spin tripling is indicated for the $N = 1$ level only. Energies are given by scale at right.

gles. The other two CO_2 modes, and all three H_2O modes, are nondegenerate.

Molecular spectra. The frequencies $c\nu$ of electromagnetic spectra obey the Einstein-Bohr equation

$$hc\nu = E' - E'' \quad (4)$$

The quantities ν, in waves per centimeter, or wave numbers (cm^{-1}), will hereafter be called frequencies, as is usual in spectroscopy, although properly only the $c\nu$ are frequencies. Molecular emission spectra accompany jumps in energy from higher down to lower levels; absorption spectra, jumps from lower up to higher levels. Both E' and E'' can be either discrete or continuous levels. If both are discrete, they give a spectrum of discrete frequencies; otherwise, they give a continuous spectrum. Discrete spectra are the main type considered here. Discrete frequencies are usually called spectrum lines because of their appearance when recorded by an optical spectrograph.

Besides its frequency, the intensity and width of a spectrum line are important. Intensities vary over wide ranges. In the extreme case of nearly zero intensity for a spectroscopic transition, the transition is called forbidden. Only a small minority of all pairs of levels yield allowed transitions. These are governed by selection rules derivable from quantum theory. See SELECTION RULES (PHYSICS).

Under disturbing influences, however, some lines are seen, weakly, which violate these rules. Further, the usual selection rules are electric dipole rules, and additional transitions become very weakly allowed if magnetic dipole, electric quadrupole, and other selection rules are also considered. The following discussion is confined to spectra which obey the electric dipole rules.

Molecular spectra can be classified as fine-structure or low-frequency spectra, rotation spectra, vibration-rotation spectra, and electronic spectra. Low-frequency spectra are discussed elsewhere. See ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY; MAGNETIC RESONANCE; MICROWAVE SPECTROSCOPY; MOLECULAR BEAMS; SPECTROSCOPY.

Pure rotation spectra. Transitions between energy levels differing only in rotational state give rise to pure rotation spectra. For diatomic molecules in $^1\Sigma$ states, Eq. (3), the relation is

$$hc\nu = E'_r - E''_r = hcB_v[J'(J'+1) - J''(J''+1)] + \cdots \quad (5a)$$

The transitions obey the selection rule $\Delta J = 1$ (ΔJ means $J' - J''$). Putting $J' = J'' + 1$,

$$\nu = 2B_v(J''+1) + \cdots \quad (5b)$$

Equation (5*b*) represents a sequence of lines spaced almost equidistantly ($2B_v$, $4B_v$, $6B_v$, . . .), and lying in the far infrared or (for small B or low J'') the microwave region. Their intensities are proportional to μ_e^2, where μ_e is the electric moment at R_e (see Fig. 2); hence homopolar molecules (H_2, N_2, etc.) show no pure rotation spectra. The intensities are proportional also to the lower-state (v'', J'') level population and to ν (for absorption) or ν^4 (for emission).

Pure rotation spectra of linear polyatomic molecules are like those of diatomic molecules. Poly-

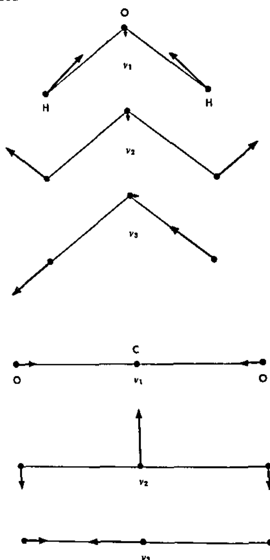


Fig 6 Normal vibration modes of H_2O and CO_2. Synchronized displacements of atoms occur in proportion to lengths of the arrows. Diagram corresponds to snapshot taken at one phase of vibration.

atomic molecules having $\mu_e = 0$, whatever their shape (examples are CO_2, CH_4, C_6H_6), have no pure rotation spectra. In other cases, the spectra can be obtained using $hc\nu = E'_r - E''_r$ with appropriate E_r expressions and selection rules.

Vibration-rotation bands. Spectra involving only vibrational and rotational state changes lie mainly in the infrared. For a $^1\Sigma$ diatomic state, using Eqs. (1), (3), and (4),

$$\nu = \nu_0(v',v'') + [B'J'(J'+1) - B''J''(J''+1)] + \cdots \quad (6)$$

with

$$\nu_0 = \omega_e(1 - x_e)(v' - v'') - x_e\omega_e(v'^2 - v''^2) \quad (7)$$

Here B' and B'' mean B_v for v' and v'', respectively. Each band consists of two sets of rotational lines, one on each side of its ν_0. Each line corresponds to a particular rotational transition conforming to $\Delta J = \pm 1$. The two series (branches) have frequencies as follows: R or positive branch ($J' = J'' + 1$),

$$\nu = \nu_0 + 2B''(J''+1) + (B' - B'')(J''+1)(J''+2) \quad (8a)$$

P or negative branch ($J' = J'' - 1$),

$$\nu = \nu_0 - 2B''J + (B' - B'')J''(J'' - 1) \quad (8b)$$

Both can be represented by a single equation

$$\nu = \nu_0 + (B' + B'')M + (B' - B'')M^2 + \cdots \quad (9)$$

by letting $M = J'' + 1$ for the R and $M = -J''$ for the P branch. Neglecting the term in M^2, Eq. (9) represents a series of equidistant lines with one missing ($M = 0$) at ν_0. Figure 7 shows how the line positions are related to the upper (v') and lower (v'') sets of rotational levels.

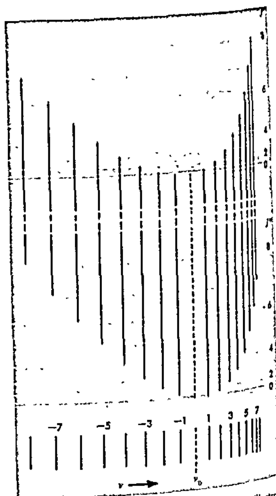
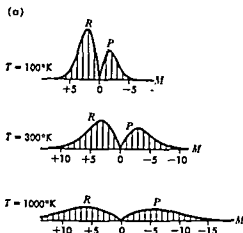


Fig. 7. Relation of band lines (lower part of illustration), see Eqs. (8) and (9), to rotational levels, see Eq (6), for a vibration-rotation band or an electronic band. In the former case, the upper and lower sets of rotational levels belong to two vibrational levels of a $^1\Sigma$ electronic state. In the latter case, they belong to two different $^1\Sigma$ states. Positive M values, R branch; negative M values, P branch.



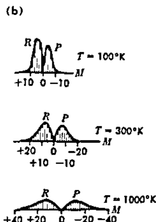


Fig 8. Intensity distribution at several temperatures for a diatomic absorption band. Line positions are based on Eq. (9) assuming $B' = B''$ for simplicity; frequency increases toward the left (opposite to Fig. 7) (a) and (b) correspond respectively to B values of HCl ($B = 10.44$ cm^{-1}) and of 2 cm^{-1} (approximately the value for CO, for which $B = 1.93$ cm^{-1}). (After G Herzberg, *Molecular Spectra and Molecular Structure*, vol 1, 2d ed, Van Nostrand, 1950)

Since $B' - B''$ is a small negative quantity [see Eq. (3a), noting that $v' > v''$], the M^2 term makes the P line spacing increase and the R line spacing decrease slowly as M increases. This is shown, exaggerated, in Fig. 7. At some large M value, the R branch turns back on itself, but usually the lines have become weak before this value is reached.

The relative intensities of band lines depend primarily on the initial rotational distribution of molecules. More precisely,

$$\text{Intensity} = C(v',v'')\nu^n(J'+J''+1)e^{-B_{in}J_{in}(J_{in}+1)hc/kT} \tag{10}$$

Here B_{in}, J_{in}, and n in ν^n are B', J', and 4, respectively, for an emission band, and B'', J'', and 1, respectively, for an absorption band. Figure 8 shows diagrammatically how the values of B and T affect the appearance of a typical absorption band ($B' = B''$ has been assumed for simplicity in Fig. 8) Figure 9 shows the appearance of an actual HCl band. The weaker HCl^{37} lines are at slightly lower frequencies than the HCl^{35} lines, mainly because ω_e is smaller [see Eqs. (1a) and (7)].

The factor $C(v',v'')$ is largest by far for fundamental bands ($\Delta v = 1$), and falls rapidly with increasing Δv in the overtone bands or harmonics (Δv is $v' - v''$). For fundamental bands, C depends on the slope of the $\mu(R)$ curve (see Fig. 2), being approximately proportional to $(d\mu/dR)^2$ taken at R_e. For overtone bands, C depends on the

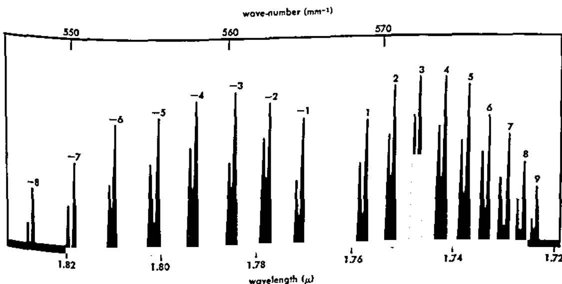


Fig 9. First harmonic (2,0) vibration-rotation band of HCl in absorption. R branch to right, P branch to left, showing intensity distribution. The stronger lines are HCl^{35}, the weaker companions HCl^{37}. (After C. F. Meyer and A. A. Levin, *Phys. Rev.*, 34·44, 1929)

detailed shapes of both the $\mu(R)$ and $U(R)$ curves. Fundamental or *overtone bands arising* from $v'' > 0$ are called hot bands.

Vibration-rotation absorption bands of liquids and solutions are widely used in chemical *analysis.* Here the rotational structure is *blurred out,* and only an "envelope" is seen. For many purposes, it is sufficient to know *empirically the spectrum* of each molecule which *may be present.* Also, groups of atoms which recur in many molecules often have approximately constant *frequencies which* can be used for identification and *in determining* molecular structures. *See* INFRARED SPECTROSCOPY.

Electronic band spectra. These are the most general type of molecular *spectra.* The characteristic feature is a change of electronic state. From Eqs. (2) and (4), neglecting fine structure,

$$\nu = \frac{(E'_{el} - E''_{el}) + (E'_v - E''_v) + (E''_r - E''_r)}{hc} \quad (11a)$$

$$\nu = \nu_{el} + \nu_v + \nu_r = \nu_0 + \nu_r \quad (11b)$$

Diatomic electronic spectra are often observed in emission, while the electronic spectra of polyatomic molecules are *usually* absorption spectra. Depending mainly on the magnitude of ν_{el}, electronic spectra occur *in the* infrared, visible, ultraviolet, or vacuum ultraviolet.

For any one electronic transition, the spectrum consists *typically* of many bands. These lie in general at *frequencies* both above and below ν_{el}, since ν_v can be positive or negative. They constitute a band *system.* Each band consists of numerous rotational lines arranged in two or more branches and lying on both sides of a central position ν_0.

For *diatomic* molecules, ν_0 depends on a *single* v' and v'' *and,* using Eq. (1) for each electronic state, is given by

$$\nu_0(v',v'') = \nu_{el} + [\omega'_e(v' + \tfrac{1}{2}) - x'_e\omega'_e(v' + \tfrac{1}{2})^2 + \cdots] - [\omega''_e(v'' + \tfrac{1}{2}) - x''_e\omega''_e(v'' + \tfrac{1}{2})^2 + \cdots] \quad (12)$$

Since ω_e and $x_e\omega_e$ are now different (often strongly) in the upper and lower states, $\nu_0(v',v'')$ cannot be reduced to as simple an expression as the corresponding Eq. (7) for vibration-rotation bands. Eq. (12) is more convenient when rewritten as

$$\nu_0(v',v'') = \nu_{00} + (\omega'_0 v' - a'v'^2) - (\omega''_0 v'' - a''v''^2) + \cdots \quad (12a)$$

where

$$\nu_{00} = \nu_{el} + \tfrac{1}{2}(\omega'_e - \omega''_e) - \tfrac{1}{4}(x'_e\omega'_e - x''_e\omega''_e)$$
$$\omega'_0 = \omega'_e(1 - x'_e) \qquad a' = x'_e\omega'_e \quad \text{etc.}$$

The *relative* intensities of the bands depend on (1) the *initial* distribution of molecules among vibrational levels, and (2) the relative *transition* probabilities *from* any initial to various *final vibrational* levels.

The simplest [illegible]
of a co[illegible]
all mol-[illegible]
[illegible] spectrum

then consists of one "v' progression," a single series of bands with various values of v'; the frequencies are given by $\nu = \nu_{00} + \omega_0 v' - a'v'^2$. For a hot or a heavy gas, additional weaker v' progressions with $v'' > 0$ also appear.

In emission spectra, the initial population usually ranges over a number of v' values, from each of which transitions occur to a number of v'' values, so that the system contains many bands on both sides of ν_{00}. In the special case of fluorescence spectra, the molecule is excited to various v' values by absorbing light; it then emits light belonging to the same (or sometimes another) electronic transition. From each v', it can descend not only to the original v'', but also to various other, mainly larger, values. Hence, fluorescence bands lie mainly at lower frequencies than the absorption bands used to excite them.

Relative transition probabilities are governed by the Franck-Condon principle. This takes note of the very great rapidity of electronic motions as compared with those of the far more massive nuclei, and concludes that during the extremely brief time for an electronic transition, the nuclei tend to remain unchanged in their positions and momenta. It is applicable to both polyatomic and diatomic spectra. Consi[illegible] starting from the [illegible] $v'' =$ [illegible] lower [illegible] the bottom point A ($v'' = 0$ if zero-point vibration is neglected) to point B on any one of the upper curves would correspond to an electronic absorption transition in which the nuclei have not moved.

In case (a) of Fig. 10, point B corresponds to $v' = 0$, and the conclusion is that this is the most probable v' for $v'' = 0$. In case (b), point B cor-

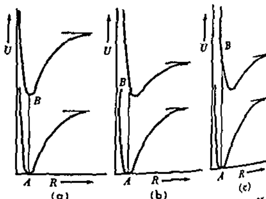


Fig. 10. Diatomic $U(R)$ curves for three cases to explain the vibrational intensity distribution according to the Franck-Condon principle. The asymptote of each curve for large R corresponds to dissociation into atoms, with one or both atoms excited in the case of the upper curves. Starting in each case from the bottom of the lower curve (essentially $v'' = 0$), the most probable transition in absorption is to $v' = 0$ in (a), $v' = 3$ or 4 in (b), and to the dissociation continuum in (c), as indicated by the vertical lines in each case (*After G. Herzberg, Molecular Spectra and Molecular Structure, vol. 1, 2d ed., Van Nostrand, 1950*)

responds to an excited molecule at the inner turning point of a vibration with a v' of possibly 3 or 4, in a typical case. One then concludes (with J. Franck) that the strongest absorption bands for $v'' = 0$ have $v', v'' = 3, 0$ and $4, 0$. To obtain more exact information, a quantum-mechanical calculation (first carried out by E. U. Condon) is necessary.

In case (c) of Fig. 10, point B corresponds to an energy level in the dissociation continuum above the asymptote of the upper $U(R)$ curve. According to the Franck-Condon principle, the absorption spectrum will have maximum intensity in a continuous range of frequencies with $hc\nu$ about equal to the energy difference AB. The quantum-mechanical calculation shows that the actual spectrum will extend with appreciable intensity over a range of both higher and lower frequencies than this, including, on the lower-frequency side, a number of high-v' bands. Actual examples of such spectra (a long v' progression followed by a strong continuum) are the far ultraviolet Schumann-Runge bands of oxygen and the visible bands of iodine. By measuring the frequency at which the continuum begins, one obtains an exact value of the dissociation energy of each of these molecules. In so doing, any excitation energy of the atomic dissociation products to which the upper $U(R)$ curve leads is subtracted.

The Franck-Condon method is useful in understanding intensity distributions and structure in emission as well as absorption band systems. For diatomic spectra, various patterns of intensity as functions of v' and v'' occur, depending largely on the R_e values of the two $U(R)$ curves and, of course, also on the initial distribution among v' levels. Sometimes the upper-state $U(R)$ curve is stable (has a minimum) but the lower state is repulsive. Continuous emission spectra then occur, with the atoms flying apart on reaching the lower state. The H_2 molecule shows such a spectrum, as do rare gas molecules such as He_2 and Kr_2, which are stable only in excited or ionized states.

Molecular electronic states. Before discussing the structures of electronic bands, one must consider the nature of molecular electronic states. Each electronic state has orbital and spin characteristics. The spin quantum number S has a whole-number value if the number of electrons is even, a half integral value if it is odd. Electronic states with $S = 0$ are called singlet states, all others multiplet states. The orbital characteristics differ sharply for linear (including diatomic) and nonlinear molecules.

For linear molecules only, there is a quantum number Λ such that $\pm\Lambda h/2\pi$ is the component of angular momentum around the line of nuclear centers. Linear-molecule electronic states can be discussed under three headings: (1) singlet states; (2) multiplet states with strong spin coupling (Hund's case a); and (3) multiplet states with weak spin coupling (Hund's case b). Strictly speaking, actual multiplet states are intermediate between cases a and b, or between these and certain other cases called c and d. The discussion to follow is largely restricted to singlet electronic states.

Singlet states with $\Lambda = 0$ include $^1\Sigma^+$ and $^1\Sigma^-$ states; states with $\Lambda = 1, 2, \ldots$ are called $^1\Pi$, $^1\Delta$, and so on. In linear molecules with a center of symmetry (H_2, CO_2, etc.), one must further distinguish even and odd (g and u) states: $^1\Sigma^+_g$, $^1\Sigma^+_u$, $^1\Sigma^-_g$, $^1\Sigma^-_u$, $^1\Pi_u$, $^1\Pi_g$, $^1\Delta_g$, $^1\Delta_u$, etc. The rotational levels of singlet states obey the symmetric rotor formula,

$$E_r = hc[BJ(J+1) - \Lambda^2] + \cdots \qquad (13)$$

In Eq. (13), J is restricted to integral values equal to or greater than Λ.

For $\Lambda > 0$, each rotational level is a narrow doublet (Λ-doubling). Corresponding fine structure [see E_{fs} in Eq. (2)] can usually be detected in electronic bands, although (but for ground states only) it can be much more accurately studied in low-frequency spectra. Hyperfine structure [see E_{hfs} in Eq. (2)] is usually on too small a scale to be detected in electronic band lines, but has been found in a few cases. It is best studied in low-frequency spectra.

Electronic band structures. The simplest electronic bands occur for transitions between singlet electronic states. The possible types of electronic transitions are limited by the selection rule $\Delta\Lambda = 0, \pm1$. The structures of $^1\Sigma - {}^1\Sigma$ bands are essentially the same as for the $^1\Sigma$-state vibration-rotation bands described earlier. Equations (8) to (10) and Fig. 7, also Fig. 8, for the intensities in absorption are still applicable if Eq. (12) instead of Eq. (7) is used for ν_0, and it is recognized that B' and B'' now belong to two different electronic states.

The quantity $B' - B''$ in Eq. (9), instead of always being a small negative quantity, may now be either positive or negative, and $(B' - B'')/(B' + B'')$ is often fairly large (although it can also be nearly zero). As a result, it is usual in electronic bands to find so-called heads. A head is a position of maximum or minimum frequency; by using Eq. (9) to obtain $d\nu/dM = 0$, one finds $M_{head} = (B' + B'')/2(B' - B'')$. Then, on inserting M_{head} into Eq. (9), one obtains $\nu_{head} = \nu_0 - (B' + B'')^2/4(B' - B'')$. [Since $(B' + B'')/2(B' - B'')$ is not usually a whole number, the actual M_{head} is the nearest whole-number M to that calculated.] According as $B' - B''$ is negative or positive, the positive (R) or the negative (P) branch forms the head. Figure 7, if continued to somewhat larger M values, would illustrate the formation of an R-branch head at a calculated M of 10.5; the actual head would be formed by the two coincident lines $M = 10$ and 11.

Although homopolar molecules (H_2, N_2, etc.) have no pure rotation or vibration-rotation spectra, they do have electronic spectra. For homonuclear homopolar molecules, the band lines show alternating intensities. The lines in each branch are alternately stronger and weaker as M increases, this effect being superposed on the otherwise smoothly

varying intensity distribution The alternation ratio depends on the nuclear spin I and has been, in several cases, the means of determining I. When $I = 0$, alternate lines are completely missing. Heteronuclear molecules, even if homopolar (for example, HD or $O^{16}O^{18}$) do not show alternating intensities.

Polyatomic electronic spectra. These differ from diatomic electronic spectra because several initial and final vibration quantum numbers are involved, and because the rotational structure (except for linear molecules) is usually much more complicated However, the detailed structures of the electronic spectra of a number of simple molecules and radicals in the vapor state in emission and in absorption have been studied. Nevertheless, for the most part, the spectra of polyatomic molecules are examined as absorption spectra in solution. The rotational structure is then completely blurred out, but the vibrational structure can be seen.

The Franck-Condon principle is here a useful guide. One of its corollaries, which amounts almost to a selection rule, is that only totally symmetric vibrations (vibrations during which the equilibrium symmetry of the molecule is preserved) undergo quantum number changes This greatly simplifies the vibrational structure, especially of absorption spectra where most molecules are initially mainly in the $v'' = 0$ state of all vibrations. One finds then mostly v' progressions of one or a very few totally symmetric vibrations, and combinations of these.

Rather often, polyatomic band systems do not even show obvious vibrational structure. This can happen for any of several reasons: the upper state may involve dissociation, for example, in CH_3I, the first ultraviolet absorption region yields $CH_3 + I$; there may be so many low-frequency, upper-state vibrations that the spectrum looks continuous; or there may be a combination of these and other reasons. Such continuous or pseudocontinuous band systems are often loosely referred to as bands. For complicated molecules, the spectra of several different electronic transitions often overlap strongly so that it is difficult even to separate one system from another. See ATOMIC STRUCTURE AND SPECTRA; ELECTRON SPIN; INTERMOLECULAR FORCES; MOLECULAR ASSOCIATION; MOLECULAR WEIGHT; RAMAN EFFECT, RESONANCE (MOLECULAR STRUCTURE); SCATTERING EXPERIMENTS, ATOMIC AND MOLECULAR; VALENCE.

[R.S.MU.]

Bibliography: G. Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules*, vol. 2, 2d ed., 1945; G Herzberg, *Molecular Spectra and Molecular Structure*, vol. 1, 2d ed., 1950; R. S. Mulliken, Report on notation for the spectra of polyatomic molecules, *J. Chem. Phys.*, 23:1997, 1955.

Molecular weight

The sum of the atomic weights of all the atoms in a molecule. Atomic weights (and therefore molecular weights) are relative weights, arbitrarily referred to an assigned atomic weight for oxygen. See ATOMIC WEIGHT.

Gram-mole. The amount of a substance which has a weight in grams equal to the molecular weight of the substance is called a gram-mole or a gram molecular weight. The number of molecules in a gram-mole is 6.02×10^{23} and is called Avogadro's number. Except for substances composed of only one isotope of each of their elements, division of gram-molecular weight by Avogadro's number does not give the exact weight of any molecule. For example, there are three kinds of chlorine molecules in ordinary chlorine, $Cl^{35}Cl^{35}$, $Cl^{35}Cl^{37}$, and $Cl^{37}Cl^{37}$, and the masses of the two kinds of chlorine atoms are 34.97990 and 36.97754, referred to O^{16} as 16 exactly. For additional discussion, see AVOGADRO NUMBER.

Formula weight. The ultimate analysis of a substance provides precise knowledge of formula weight, namely, the sum of the atomic weights of the atoms in the simplest formula for the sub- [illegible] multiple [illegible] formula weight, the sum of the atomic weights of carbon and hydrogen, is 13.02. Each benzene molecule is a hexagonal structure of true formula C_6H_6 and molecular weight 78.11. Some substances, such as sodium chloride, contain no discrete molecules and are best assigned formula weights but no molecular weights. Known molecules range in molecular weight from two (hydrogen) to several million (large virus molecules).

Experimental methods. Molecular weights are determined by a variety of methods, the choice in a particular instance depending on the properties of the substance, the purpose, and the facilities at hand. Historically, molecular-weight determinations were essential in establishing the stoichiometry which forms the basis for the fundamental laws of chemistry; today, in addition to their usefulness in pure science, they are made to interpret chemical reactions, to aid in determining molecular structure and molecular shapes, to give a better understanding of solvent-solute interaction, and to provide data for the design and control of a great variety of useful industrial processes.

Strictly speaking, the definition of molecular weight applies only to substances consisting of molecules of a single mass, but nearly all real samples of interest are mixtures of molecules of several masses. When molecular-weight determinations are made of mixtures, the average molecular weight obtained may depend on the method employed. In those determinations in which every molecule present produces the same effect, regardless of the size of the molecule, the result is said to be a number average.

Except where noted, the methods described here yield number average molecular weights.

Gas-density method. Equal volumes of different gases at the same (low) pressure and temperature contain the same number of molecules (Avogadro's law), and therefore their weights are proportional

to their molecular weights. From a precise standpoint the gas-density method of determining molecular weights consists in determining density as a function of pressure at constant temperature. The ratio of density, d, to pressure, p, is a linear function of pressure, and, at low pressures, every real gas approaches ideal behavior. The ideal or perfect gas equation may be written

$$M = \left(\frac{w}{v}\right)\frac{RT}{p} = \frac{d}{p}RT$$

where M is molecular weight, w weight, v volume, T absolute temperature in degrees Kelvin, R the molar gas constant, d density, and p pressure. Values of $(d/p)_{p=0}$ are equal to M/RT. The perfect-gas equation is a useful approximation at atmospheric pressure, but the pressure coefficients of d/p may be quite large and are different for different gases. At 0°C, $d/p = 2.8580 + 0.0686p$ for sulfur dioxide, while the value for oxygen is given as $d/p = 1.4276 + 0.0013p$. See GAS.

The gas-density method may be applied to easily volatile liquids. An approximate value, found in a measurement at atmospheric pressure, often is sufficiently accurate to enable a choice between a formula weight, obtained by precise chemical analysis, and some multiple of it.

Gaseous effusion. Equal volumes of different gases at the same temperature and pressure possess the same kinetic energy of molecular motion; hence, the mean velocities of the molecules of the gases are inversely proportional to the square roots of their densities and thus of their molecular weights (see KINETIC THEORY OF MATTER). The effusion method of molecular-weight determination involves passing gases through a small hole and measuring the time required for equal volumes of gases, of known and unknown molecular weights, respectively, to pass under the same conditions of pressure and temperature. If u_1 and u_2 are the root-mean-square velocities of the molecules of the two gases which have densities d_1 and d_2 and molecular weights M_1 and M_2, then the ratio of the times t_1 and t_2 taken for each gas is given by

$$\frac{t_1}{t_2} = \frac{u_2}{u_1} = \sqrt{\frac{d_1}{d_2}} = \sqrt{\frac{M_1}{M_2}}$$

The effusion method is especially valuable where only small amounts of sample are available.

Vapor-pressure lowering. The vapor pressure of a solvent at constant temperature is decreased by the introduction of a nonvolatile solute in accordance with Raoult's law, which may be written

$$p = p_0\frac{N}{n + N}$$

where n and N represent the number of moles of solute and solvent, respectively, in the solution, and p_0 and p denote the vapor pressures of solvent and solution. For dilute solutions, where Raoult's law is applicable, n is negligible in comparison with N, whence

$$\frac{p_0 - p}{p_0} = \frac{n}{N} = \frac{g/m}{G/M}$$

where g and G are the weights of solute of molecular weight m and solvent of molecular weight M. Both static methods and dynamic or air-saturation methods may be employed to determine molecular weights by vapor pressure lowering, but there are several sources of serious error. See SOLUTION.

Boiling and freezing point methods. Both the elevation of the boiling point and the lowering of the freezing point of a solvent occasioned by introduction of a nonvolatile solute are manifestations of vapor-pressure lowering, and either may be more satisfactorily measured than vapor-pressure lowering itself. In dilute solutions, the magnitude of ΔT_b, the elevation of the boiling point, and ΔT_f, the lowering of the freezing point, are proportional to the molal concentration of the solute. It can be shown that, as a useful approximation,

$$K_b = \frac{RT_b^2}{L_bN}$$

and

$$K_f = \frac{RT_f^2}{L_fN}$$

where K_b and K_f are the so-called ebullioscopic constant and the cryoscopic constant, respectively, of a solvent having a boiling point T_b, a freezing point T_f, a heat of vaporization L_b, and a heat of fusion L_f. R is the molar gas constant, N the number of moles of solute in 1000 grams of solvent, and the temperatures are in degrees Kelvin. K_b and K_f are also called the molal elevation of the boiling point and the molal depression of the freezing point. They are the values of ΔT_b and ΔT_f calculated for 1 mole of solute dissolved in 1000 grams of solvent but applicable only for much more dilute solutions. It follows as a useful approximation for dilute solutions that

$$M = K_b\left(\frac{1000}{G}\right)\frac{g}{\Delta T_b} = K_f\left(\frac{1000}{G}\right)\frac{g}{\Delta T_f}$$

where M is the molecular weight of the solute, and g is the weight in grams of solute dissolved in G grams of solvent. The boiling point T_b varies significantly with pressure changes occurring in the laboratory atmosphere. Hence ΔT_b, the difference between the boiling point of the solution and the pure solvent, is useful only when the two boiling points are referred to the same pressure.

Osmotic pressure. Though the osmotic pressure, π, of a dilute solution is given by $\pi V = nRT$, an equation identical in form with the perfect gas equation, it can be shown that the osmotic pressure equation is related to Raoult's law. V is the volume of the solution, n the number of moles of solute, R the molar gas constant, and T the absolute temperature in degrees Kelvin. The molecular weight M is given by

$$M = \frac{WRT}{\pi V} = \frac{RT}{\pi/C}$$

where C is the concentration of solute in grams per unit volume. But for nonideal systems such as those found when dealing with high molecular weight polymers, π/C may be a nonlinear function of C. In osmotic pressure measurement membranes impermeable to the solute are used. See OSMOSIS.

Monomolecular surface films The equation of state for ideal "gaseous" films is given by

$$FA = nRT$$

where F is surface pressure, A is the film area, n is the number of moles of film spread, T is the absolute temperature in degrees Kelvin, and R is the molar gas constant in ergs per mole per degree. Pressure-versus-area determinations are made with a film balance, and a plot of FA versus F, extrapolated to $F = 0$, gives the value of nRT, corrected for the area occupied by the molecules themselves and fully appropriate for calculation of molecular weight. Unfortunately the method has often been applied to coherent films and the results are consequently spurious.

X-ray methods. The volume occupied by a molecule in a crystal may be determined by means of x-rays. Multiplication of this volume by the measured density of the crystal and by Avogadro's number gives the molecular weight. The method is especially useful for determining the chemical formulas of complex crystalline substances, for example, ammoni[illegible]
LOGRAP[illegible]
Mass [illegible]
determ[illegible] with the mass spectrograph. For an element such as chlorine, consisting of isotopes, the chemical molecular weight is found by determining the physical atomic weights of the different isotopes and also their relative abundance. See MASS SPECTROSCOPE.

Electron microscopy. When large molecules may be isol[illegible]
th[illegible]
re[illegible]
TE[illegible]

Viscosity measurements. The magnitude of the viscosities of solutions or melts of polymeric substances may be correlated with the molar concentration of the polymeric substances in the solutions or melts. In consequence, viscosity measurements may be used in certain instances to determine molecular weights. The results have been called viscosity-average molecular weights. The method is most common in polymer science and technology, because of the ease in performing the measurements and the simplicity of the equipment. The[illegible]
[illegible]

Ultracentrifugation. The molecular weights of large molecules have been determined from sedimentation studies of solutions subjected to intense gravitational fields. In the sedimentation equilibrium method the solution is centrifuged until an equilibrium is established between sedimentation and diffusion; sometimes this requires several weeks. In the sedimentation rate method, determination is made of the rate of the diffusion of solute molecules in solution into pure solvent in contact with the solution. Concentrations are determined by optical methods (measurement of absorption and refractive index). The equations for calculation of molecular weight are applicable only for dilute solutions. An extrapolated value for the molecular weight, M, found at zero concentration by plotting $1/M$ versus concentration, is often reported. Sedimentation studies give so-called weight [illegible]
ployed. See ULTRACENTRIFUGE.

Light scattering. When a beam of light induces electronic transitions in a material, the material serves as a secondary source of light and emits scattered radiation. See SCATTERING (ELECTROMAGNETIC RADIATION). Molecular weights of polymeric materials in solution have been determined by methods involving determination of light scattering. Weight-average molecular-weight values are obtained.

Velocity of sound. The molecular weight of some pure liquids has been correlated with the velocity of sound in the liquid. The molar sound velocity R in these liquids is given by

$$R = \frac{M\gamma^{1/3}}{d}$$

where M is the molecular weight, γ the sound velocity at temperature T, and d the density at the same temperature T. Molar sound velocity R in these liquids is an additive property of atoms and bonds similar to molar refraction, N, which is given by

$$N = \frac{M}{d}\frac{(n^2 - 1)}{(n^2 + 2)}$$

where n is the refractive index. It has been shown that R and N are related by the formula

$$R = 41.59N + 57.7$$

for the polyethylene glycols. [J.R.A.]

Bibliography: P. J. Flory, *Principles of Polymer Chemistry*, 1953; A. Weissberger (ed.), *Technique of Organic Chemistry*, vol. 1, pt. 1, 1959.

Molecule

A molecule may be thought of either as a structure built of atoms bound together by chemical forces or as a structure in which two or more nuclei are maintained in some definite geometrical configuration by attractive forces from a surrounding swarm of negative electrons. Besides chemically stable molecules, short-lived molecular fragments called free radicals can be observed under special circum-

stances, for example at high temperatures, in electrical discharges, in chemical reactions, and even (but in small quantities) frozen into ordinary solid substances under some conditions. Free radicals are really just highly active molecules. *See* FREE RADICAL; *see also* CHEMICAL BINDING; MOLECULAR STRUCTURE AND SPECTRA. [R.S.MU.]

Mollusca

One of the divisions or phyla of the animal kingdom containing the snails, slugs, octopuses, squids, clams, mussels, and oysters. The Mollusca contain well over 100,000 described species. This phylum is characterized by a shell-secreting organ, the mantle, and a radula, a food-rasping organ located in the forward area of the mouth. A somewhat different type of mantle occurs in the Brachiopoda and an organ similar to the radula is known in a few annelid worms. The body of the mollusk is primarily bilaterally symmetrical, with the mouth at one end and the anus at the other, the alimentary canal forming an axis. In the snails, the Gastropoda, this bilateral symmetry is modified by torsion, that is, a twisting of the animal of 180°, so that the main nervous system forms a figure 8 and the anal opening is brought forward toward the mouth. On the ventral surface there is a muscular organ, the foot, by means of which the animal is able to crawl.

Modifications of this muscular organ are found as the tentacles and the funnel in the Cephalopoda and the plowing organ in the Pelecypoda. The pallium or mantle is attached in the dorsal area and is more or less open along the ventral area. Its forward edge and outer surface secrete the shell. The mantle cavity is the area between the two lobes of the mantle which contains the visceral mass in addition to the gills. In the Pulmonata, the gills are absent and in their place there is a pulmonary sac surrounded by blood vessels; this functions as a lung. *See* ANIMAL KINGDOM.

CLASSIFICATION

The phylum Mollusca is divided into six classes: Monoplacophora, Amphineura, Scaphopoda, Gastropoda, Pelecypoda, and Cephalopoda. See separate articles on these groups. An outline of their classification follows.

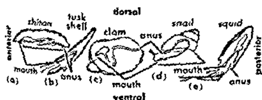


Fig. 1. Phylum Mollusca. Relations of the shell (heavy lines), foot (stippled), digestive tract (shaded), mouth and anus. (a) Amphineura. (b) Scaphopoda. (c) Pelecypoda. (d) Gastropoda. (e) Cephalopoda. (T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957)

Class Monoplacophora. Shell caplike; soft structures segmented; only two known Recent species; marine.
Class Amphineura. The chitons.
 Subclass Aplacophora. No shell; only calcareous spicules in the integument, marine.
 Subclass Monoplacophora. External shell; marine.
 Subclass Polyplacophora. Shell composed of eight transverse plates; marine.
Class Scaphopoda. The tusk shells. A tubular, univalve shell which is open at both ends; marine.
Class Gastropoda. All snails and slugs.
 Subclass Prosobranchia. Mainly operculate snails with separate sexes; marine, land and fresh water.
 Subclass Opisthobranchia. Mainly without an operculum; hermaphroditic; shell external, internal, or absent; marine and fresh-water.
 Subclass Pulmonata. Snails which breathe by means of a pulmonary sac, with few exceptions without an operculum; hermaphroditic; usually an external shell; land, fresh-water, and rarely marine.
Class Pelecypoda. Shell bivalve, rarely multivalve.
 Subclass Protobranchia. Adductor muscles about equal in size; hinge taxodont or simple; gills for respiration only; labial palps greatly developed for obtaining food; marine.
 Subclass Taxodonta. Adductor muscles about equal in size; hinge taxodont; gills for respiration and obtaining food; marine.
 Subclass Anisomyaria. Adductor muscles unequal in size or limited to one; hinge variable; gills for respiration and obtaining food; marine.
 Subclass Eulamellibranchia. Adductor muscles about equal; hinge variable; gills for respiration and feeding; marine and fresh-water.
 Subclass Septibranchia. Adductor muscles about equal; gills transformed into a muscular septum; ligament external; resilum subinternal.
Class Cephalopoda. The squids, octopuses and nautilus; marine.
 Subclass Nautiloidea. Shell external, coiled and chambered in the few recent species; coiled or straight in the fossil forms.
 Subclass Ammonoidea. Shell external, coiled and chambered; no recent forms.
 Subclass Coleoidea. Shell internal; calcareous, chitinous or completely lacking.

MORPHOLOGY

Shell. In most groups of snails the posterior and dorsal surface of the foot produces a calcareous or horny plate, the operculum. When the animal retracts within the shell, the operculum forms a plug for the aperture. The shell is composed of calcium carbonate which is secreted within the matrix of conchyolin, an organic, nitrogenous substance closely allied to chitin. The growing edge of the shell is produced by the forward edge of the mantle; the thickness of the shell is produced by

epithelial tissue on the outer side of the mantle. Most mollusks produce a shell; in others, such as certain slugs, it is reduced to a semi-internal plate, and in others it may be completely absent. In the squids, it is reduced to a horny pen along the dorsal margin.

Radula. In all classes other than the bivalves, the Pelecypoda, the mouth contains an organ called the radula. This is a filelike ribbon studded with toothlike structures. During feeding, this organ is pulled very rapidly back and forth, the minute teeth rasping off particles which are taken into the mouth. The structure of the radula varies greatly among the different groups of mollusks. Plant feeders have many teeth with small denticles, whereas animal feeders have few teeth and large hooklike denticles. In certain groups of the Gastropoda, the Toxoglossa, the radula has become much modified, the teeth have become elongated and spearlike and in addition are associated with a poison gland. The poison is exceedingly toxic and is used by these mollusks to procure food. Getting close to its prey, the snail drives one of these dartlike teeth into its victim, which dies in a very short time. The prey, a small fish or a worm, is then drawn within the proboscis and swallowed. Twenty or more human deaths are known to have occurred from careless handling of these venomous mollusks.

Shells. The shell of the bivalves is composed of two valves which are connected by an elastic hinge ligament at a point on the dorsal margin. The ligament causes the valves to open. One or two adductor muscles which extend across the body from valve to valve close the valves. In the univalves, particularly the gastropods, an adductor muscle is attached to the columella, the axis of the shell. By contraction of this muscle, the animal can be retracted within the shell. The shell is usually composed of three layers. The outer layer, the periostracum, is organic in nature and is allied to chitin. This layer serves to protect the shell from acid action. Below the periostracum, the two layers of the ostracum are composed of carbonate of lime formed within a framework of conchyolin, similar in chemical structure to the periostracum. The first of these layers is composed of prismatic cells of lime formed at right angles to the laminated layer,

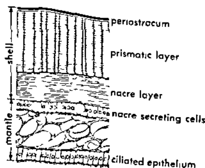


Fig. 2. Fresh-water clam. Enlarged cross section of shell and mantle. (*T. I. Storer and R. L. Usinger, General Zoology, 3d ed., McGraw-Hill, 1957*)

the inner layer. In most snails and bivalves the shell is porcelaneous in structure, the laminated layers being more or less continuous across the shell. A few families in both the gastropods and bivalves have nacreous shells. Here the laminated layer is formed of exceedingly small, thin plates, the margins forming minute prisms which reflect light. The two valves in a bivalve shell are connected on the dorsal margin with an elastic hinge ligament which causes the valves to open. The valves are closed by one or two adductor muscles which extend across the body.

Digestive system. The alimentary canal in most forms has become convoluted to increase the amount of digestive surface. Behind the gullet there is a saclike stomach which possesses ducts leading from a large digestive gland. In many mollusks, particularly those depending upon plant food, there is a rodlike structure which protrudes into the stomach. This is called a crystalline style By means of cilia, this rod rotates upon a plate which wears away particles of the style containing an enzyme. The enzyme aids in the digestion of plant material. Most of the bivalves are filter feeders; that is, small particles of food are brought in by means of the incurrent siphon. These food particles are captured by the mucus on the gills and are carried by ciliary action into the food tracts and then into the mouth.

Circulation. The heart is dorsal to the digestive tract and is enclosed in a sac, the pericardium It contains a muscular ventricle, which forces the blood through the body, and usually two auricles, which receive the blood from the gills and transmit it to the ventricle.

Reproduction. In most marine snails and bivalves the sexes are separate. The land snails contained in the subclass Pulmonata are hermaphroditic, each individual containing both sexes, although mating usually occurs between two individuals. Most marine bivalves give birth to their young at an early stage, either as eggs which shortly hatch, or as young larvae. The larvae pass through two larval forms, the trochophore and the veliger stages, both of which are free-swimming and are subject to wide dispersal by ocean currents. During the veliger stage the shell begins to form and in the late veliger stage the velar lobes or swimming organs disappear and the animals settle to the bottom. Many marine snails follow this same procedure; in others, the eggs are deposited in capsules and the two larval stages are passed within the capsule. In a few cases the capsule ruptures during the veliger stage, releasing the young for a brief free-swimming period. Most land snails lay eggs, usually in the ground in small clusters, from a few up to 300–400. A few of the land snails are oviparous; that is, they produce eggs but the young are hatched within the brood pouch and then emerge as very small snails.

Habits and ecology. The mollusks are one of the most adaptive of all animal groups. They occur in the great ocean depths, 30,000 ft (bivalves), and

have been recorded at altitudes of 18,000 ft (land pulmonates), a vertical range of more than 9 miles. They occupy all types of environment as long as there is some lime from which they can form their shells. Even areas devoid of lime may have some of the shell-less mollusks, the land slugs. Land snails are usually secretive in habit, living under

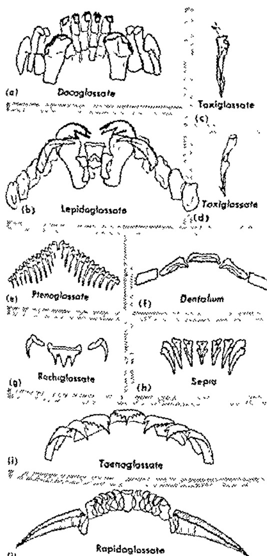


Fig. 3 Types of radula. (a) Docoglossate radula from *Ancistromesus*, an amphineuran. (b) Lepidoglossate radula from *Cryptoconchus*, a marine gastropod. (c), (d) Toxiglossate radulae from *Conus* and *Terebra*, marine gastropods. (e) Ptenoglossate type from *Actaeon*, a marine gastropod. (f) Radula of *Dentalium*, a common scaphopod. (g) Rachiglossate radula of *Murex*, a marine gastropod. (h) Two transverse rows from the radula of *Sepia*, a common sepioid squid. (i) Taenoglossate radula of *Lanistes*, a fresh-water gastropod. (j) Rapidoglossate radula of *Clanculus*, a marine gastropod. (R. R. Shrock and W. H. Twenhofel, *Principles of Invertebrate Paleontology*, 2d ed., McGraw-Hill, 1953)

leaves during the day and feeding mainly at night when the air is cool and moist. In many tropical areas divergent groups of snails live in trees, feeding on the lichens which grow on the bark. Others live only on limestone rocks, also feeding on lichens. Marine snails occupy a multitude of habitats. They exist in the spray zone, in tide pools, in mud and sand flats, on rocks and coral, and even bore into soft rocks and coral. They range from the spray zone to the profound depths. Marine bivalves occur from the upper tide level in sand and mud exposed on rocks and on mangrove roots, and like the snails, range from the upper tide level to the great ocean depths. Fresh-water snails and clams are also found in every ecological niche, from the quiet waters of small ponds and streams to rapids and swiftly flowing areas of the largest rivers. [W.J.C.]

Molluscum contagiosum

A skin disease caused by a virus and characterized by the appearance of small (average diameter, 2 mm) discrete lesions in groups on the face, arms, or genitalia. They are firm and pearly white with a sharply indented central core and yield an infectious filtrate which produces the disease when inoculated into human volunteers. The skin lesion, which will disappear without treatment, contains large, acidophilic, granular masses known as the molluscum body which, when disrupted, releases smaller elementary bodies (Lipschutz bodies). The disease, which may be epidemic in children, occurs in all ages and is world-wide in distribution. See ANIMAL VIRUS [A.E.M.]

Molpadonia

A small order of Holothuroidea characterized by the short, plump body bearing a tail-like prolongation. Simple fingerlike or featherlike tentacles as well as respiratory trees are present. The tube-feet are restricted to a zone about the cloaca. They feed on bottom material, and live buried in the sand or mud, with only the anterior and posterior ends exposed. Although they are sedentary and seldom emerge from their burrows, Molpadonia may be dislodged and cast ashore in vast numbers by a heavy storm. The single family, Molpadiidae, comprises 12 genera with species in most seas, and is usually found at moderate depths. An example is *Caudina*. See HOLOTHUROIDEA. [H.B.F.]

Molybdate

A compound containing molybdenum in the 6+ oxidation state and derived from molybdic acid, H_2MoO_4. The normal molybdates, MoO_4^{2-}, are mainly insoluble salts. These salts are soluble in strong acid and form condensed ions or isopolymolybdates, for example, $(NH_4)_6Mo_7O_{24}\cdot 4H_2O$. Although these isopolymolybdates are complex enough, molybdates also form polymeric compounds with anhydrides of other elements, such as phosphorus, which are called heteropoly compounds. The yellow precipitate which is used in

qualitative and quantitative analysis of phosphate is this type of compound and can be written as $(NH_4)_3[P(Mo_3O_{10})_4]$.

The properties of the normal molybdates are similar to those of the sulfates, chromates, and tungstates. The only soluble normal molybdates are the salts of ammonium, sodium, potassium, rubidium, lithium, magnesium, beryllium, and thallium.

If molybdates in an acid solution are carefully reduced, a strong blue color forms. The exact composition of the compounds giving this color is not known. Molybdates are used as pigments, chemical reagents, and corrosion inhibitors. *See* MOLYBDENUM. [E.E.WR.]

Molybdenite

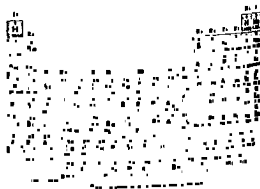
A mineral having composition MoS_2. Molybdenite is the chief ore of molybdenum. It crystallizes in the hexagonal system but crystals are rare and when found are hexagonal plates. It is commonly in scales or foliated masses. There is one direction of perfect cleavage yielding flexible but nonelastic folia. The mineral is sectile and has a greasy feel. The hardness is 1.5 (Mohs scale) and the specific gravity is 4.7. The luster is metallic and the color lead gray. Molybdenite and graphite have long been confused because of their nearly identical physical properties. They may be distinguished by the streak left on glazed paper, black for graphite and green for molybdenite. Molybdenite has been used as a lubricant.

Molybdenite is found as an accessory mineral in certain granites and pegmatites but more commonly is associated with fluorite in vein deposits of tin and tungsten. It is also present in some contact metamorphic deposits. It occurs in various places in Norway, Sweden, Australia, England, China, and Mexico. In the United States, molybdenite is found in small amounts at many localities but the most important occurrence is at Climax, Colorado, where it is found in quartz veinlets in a silicified granite with topaz and fluorite. Molybdenite as an ore of molybdenum is mined there on a large scale, making Climax the world's largest producer of molybdenum. See MOLYBDENUM. [C.S.HU.]

Molybdenum

Chemical element number 42, molybdenum, Mo, with atomic weight 95.95, was discovered by K. W. Scheele in 1778 in molybdenite, MoS_2, the chief ore of molybdenum. The gray, heavy (specific gravity 10.2), hard (Vickers penetration number 160–220), refractory (melting point $2622 \pm 10°C$) metal is useful principally in the targets of x-ray tubes and in structural members in high-vacuum tubes for electronic purposes because of the metal's ability to form tight seals with glass. It has been suggested as a structural metal in the refractory parts in high-temperature applications. Molybdenum, an important alloying element, imparts hardness and corrosion resistance to both ferrous and nonferrous alloys containing it.

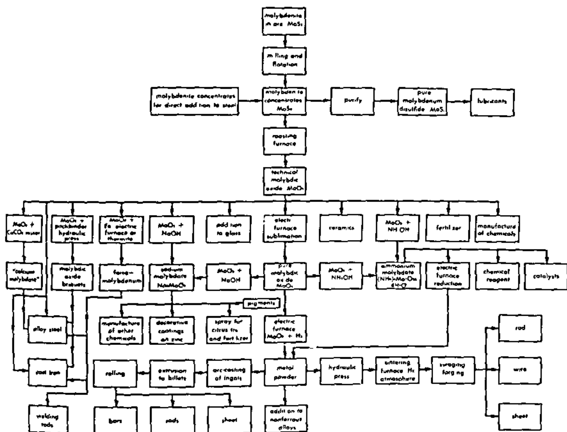
Properties. Molybdenum is a transition element of group VI of the periodic system. In its various compounds it possesses valences of 2+, 3+, 4+, 5+, and 6+, and its coordination number may be 4 or 6. Molybdenum in compounds of the lower valences readily disproportionates to yield compounds of both higher and lower valences, so that pure chemical individuals are rare in the lower valence range. Furthermore, molybdenum compounds are highly sensitive to moisture, and react readily with it through hydrolysis to oxy and hydroxy compounds or to form coordinated hydrates. All of this greatly complicates its chemistry and thus introduces uncertainties into present knowledge of it.



Occurrence. Reported consumption of molybdenum in all forms in 1957 was estimated at 60,000,000 lb, of which the United States consumed 35,000,000 lb. The approximate pattern of use of molybdenum in the United States in 1957 was ferrous alloys, 89%; metallic molybdenum (all forms), 4%; and chemical manufacture, 7%.

Fc [illegible] are known, but [illegible] powel
only [illegible] in the
lite [illegible] orked
Unit [illegible] lead
for molybdenum; it occurs along with other ores. Powellite occurs with scheelite, an important ore of tungsten, and is recovered and worked as part of tungsten operations. Molybdenite occurs in altered granite and pegmatite. Commercial molybdenite ores worked in 1957 in the United States yielded only 0.367% molybdenite. Fine grinding and flotation are the principal processes of beneficiating the ore. These produce a concentrated molybdenite that is roasted to molybdenum trioxide (MoO_3), the starting point for producing all other molybdenum compounds.

Extraction from ores. The roasting of molybdenite requires very little extraneous fuel. Hearth temperatures are closely controlled by drawing a large excess of air horizontally across the hearths. Molybdic oxide or molybdenum(VI) oxide begins to sublime at about 600°C, is quite volatile at 700°C, and melts at about 800°C. Thus, high temperatures mean both excessive volatilization losses and serious operating difficulties. The roasted concentrate from the furnaces is technical molybdic oxide.



Principal commercial forms and uses of molybdenum

Calcium molybdate can be made by mixing roasted concentrate with quicklime and heating this mixture. This method has been replaced by a simpler process of introducing lime on the hearth, after eliminating substantially all the sulfur, to effect the reaction between lime and molybdic oxide directly on the hearth. The uniformity of the product depends entirely on even feeding of the lime and uniform mixing. Uncalcined, pulverized limestone of high quality can replace lime. The limestone is mechanically mixed with roasted concentrate to give a product that can be introduced into molten iron. This mixture becomes calcium molybdate which is easily reduced by the molten metal to molybdenum.

Pure molybdic oxide is made by volatilizing the technical grade in an electric furnace having a hearth temperature of approximately 1000°C. The volatilized oxide (99.97% MoO_3) is collected in a bag filter.

Ferromolybdenum is made mainly by the thermite process. A typical thermite mix contains molybdenum trioxide, aluminum, ferrosilicon, high-grade iron ore, limestone, and high-grade fluorspar. The smelting reaction is complete in about 20 min.

Principal compounds. Chemical uses of molybdenum compounds include (1) reagents employed both in the detection and determination of phosphorus, alkaloids, uric acid, xanthine, and creatinine, and as oxidation-reduction indicators; (2) precipitants for forming color lakes (toners) from dyes; (3) inorganic molybdate pigments of special brilliance and hiding power in the orange-red color range; (4) constituents of vitreous glazes and enamels to which they impart valuable adherence and opacity; (5) compounds in electroplating baths used to give black coatings for decorative and protective purposes; (6) catalysts in which the molybdenum compound serves either as the

growth of plants, particularly legumes.

Molybdenum forms two well-defined oxides; a number of other less-certain binary oxides and hydrated oxides have been described. Reaction of the oxide of hexavalent molybdenum with oxides of other metals forms the important series of molybdates and polymolybdates; these can also form heteropolymers with anionic groups, notably the phosphomolybdates. Corresponding sulfides and thiomolybdates are known. With halogens, molybdenum forms a hexafluoride, a number of chlorides, and unstable bromides, all of which are sensitive to oxidation and hydrolysis. It forms no clearly

Molybdenum disilicide is produced in much the same way as ferromolybdenum from a thermite mix containing molybdenum trioxide, aluminum, 90%-ferrosilicon, 50%-ferrosilicon, limestone, and high-grade fluorspar.

Anhydrous sodium molybdate is manufactured by the conventional process of dissolving technical molybdic oxide in caustic soda solution, filtering, evaporating, crystallizing, and drying. Commercial ammonium molybdates include a metal grade (which is the purest compound of molybdenum commercially available) and a reagent grade. The high purity of the special metal grade of ammonium molybdate ensures high purity in the metal obtained from it by reduction by hydrogen.

Oxides. Molybdenum forms the stoichiometric oxides MoO_2 and MoO_3. Intermediate oxide phases are also known. The molybdenum oxygen systems are of more than usual complexity.

Molybdenum blue Mild reduction of an acid solution of a molybdate gives a characteristic strong blue color as the first step in the series resulting from reduction Similarly, mild oxidation of acid solutions of molybdenum compounds of the lower valences produces a strong blue color as apparently the last stage in oxidation before the final sexivalent compounds A third method of preparation is to mix solutions of quinquevalent molybdenum and of sexivalent molybdenum at pH of 4 or above.

Molybdenum blue is one or a mixture of several possible compounds of quinquevalent with sexivalent molybdenum, a molybdenyl molybdate. The ratio of molybdenum to oxygen is greater than 1:2.5, but less than 1:3. Considerable uncertainty exists concerning the exact composition of the blue material, and there is good reason to believe that it is always a mixture of several compounds of such sensitivity to oxidation-reduction that they can [illegible]

[illegible] blue color formed particularly from the reduction of phosphomolybdic acid as the basis of colorimetric methods of analysis and as an indicator in other oxidation-reduction reactions.

Sulfides. Molybdenum forms a series of sulfur compounds closely corresponding to the oxides and including a further series in which sulfur simply replaces oxygen in various oxysulfides of molybdenum and in the thiomolybdates. Four molybdenum sulfides have been described: the sesquisulfide, Mo_2S_3; the disulfide, MoS_2; the pentasulfide (dimolybdenum pentasulfide), Mo_2S_5; and the trisulfide, MoS_3, corresponding to molybdic anhydride.

By far the most important of these is the disulfide (MoS_2), molybdenite, an important lubricant. Interest in the other sulfides, as well as in the thiomolybdates, is primarily academic, although they provide the means of separating molybdenum from other metals in chemical analysis and in certain industrial purifications.

In the usual analytical procedure, molybdenum, if present, is precipitated as the trisulfide (MoS_3) from acid solution by hydrogen sulfide. Precipitated molybdenum trisulfide readily dissolves in ammonium sulfide solution to form ammonium thiomolybdate. This is a general method for preparing the thiomolybdates.

Under ordinary conditions, molybdenum disulfide (molybdenite) is quite stable, but it is subject to oxidation, and weathers to a hydrated molybdenum oxide. It crystallizes in the hexagonal system with two molecules in the unit cell. Molybdenite so closely resembles graphite in its gross properties that the two have long been confused. Upon this group of physical properties resembling those of graphite depends the use of molybdenite as [illegible]

converts it to $MoCl_5$; however, bromine hardly attacks the mineral at all. Aqua regia dissolves MoS_2, and the mineral is oxidized to MoO_3 by hot concentrated H_2SO_4. Hot HNO_3 also effects an oxidation. Other acids do not affect the mineral.

Halides. Molybdenum forms compounds of widely different degrees of stability with the halogens. The binary halides have been reported as being insoluble in water, but are so sensitive to it that they react on contact to yield oxy and hydroxy halides. Some of these oxy halides are soluble and ionize to form binary cations containing molybdenum and oxygen; or they may form complex anions containing molybdenum and halogen atoms; or they may yield ternary anions or cations composed of molybdenum, oxygen, and halogen atoms.

The only binary compound with sexivalent molybdenum formed by any of the halogens is molybdenum hexafluoride (MoF_6). It is doubtful that there are binary compounds of fluorine with molybdenum of lower valences. Molybdenum is quinquevalent in the compound formed by the action of an excess of chlorine on molybdenum ($MoCl_5$), and other binary chlorides are said to exist in which molybdenum has valences of 4, 3, and 2, but not 6. The tetrabromide represents the highest valence state of molybdenum in binary combination with bromine, and bromides of both trivalent and bivalent molybdenum are also reported. Considerable doubt surrounds binary molybdenum iodides; reports of the existence of both di- and tetraiodides have been contradicted, and no others have been described.

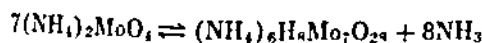
Both the hexafluoride and the pentachloride received special attention during World War II, the first in relation to isotope-separation processes, and the second as the intermediate for making molybdenum hexacarbonyl, used as a means of depositing molybdenum mirrors.

Molybdates. Molybdic acid forms a series of stable normal salts of the type, Na_2MoO_4. Generally, the normal molybdates are insoluble in water, with the exceptions of the salts of am-

monium, sodium, potassium, rubidium, lithium, magnesium, beryllium, and thallium. Acids readily dissolve the insoluble salts, and presumably form polymolybdate ions in solution.

In general, the normal molybdates resemble the sulfates and, even more closely, the normal chromates and tungstates. Like these families of salts, each soluble molybdate commonly forms several hydrates, and solutions of the soluble molybdates readily dissolve MoO_3 to form solutions from which isopolymolybdates crystallize. The several series of isopolymolybdates extend in some cases to a 16-molybdic acid salt, but not all members of any of these series are known with equal certainty. Compositions of the higher members of the series are not accurately known and probably cannot be because of the complexity of the medium in which they are formed. In contrast to the extraordinary number and variety of acid molybdates, basic molybdates are extremely rare, the basic lead molybdate, $PbO \cdot PbMoO_4$ being the only compound in this class that is well authenticated. Lead molybdate has a special interest because of its industrial uses in molybdenum orange pigment and in ceramic glazes.

Normal ammonium molybdate readily loses ammonia and becomes the paramolybdate



$$7(NH_4)_2MoO_4 \rightleftharpoons (NH_4)_6H_4Mo_7O_{24} + 8NH_3$$

This reaction is reversible, and the various ammonium molybdates encountered represent its various stages. The MoO_3 content of ammonium paramolybdate (81.53%) is high, and its solubility in water is much greater than that of MoO_3. Hence, ammonium paramolybdate with added MoO_3 (commonly designated molybdic acid 85%) can often be used as a more convenient source of molybdate ions than MoO_3 alone.

Acid solutions of molybdates react with hydrogen peroxide to yield highly colored (yellow to orange) peroxymolybdates, and these are in turn strong oxidizing agents. Two types of peroxymolybdates are known, corresponding to H_2MoO_5 and H_2MoO_8.

Molybdates readily form heteropoly compounds as well as the isopolymers noted above. These compounds are formed in acid solution by the condensation of molybdate groups around a central atom or group that may be phosphorus, arsenic, silicon, or any of a number of others. Most important of these compounds are phosphomolybdic acid and the phosphomolybdates.

Both cyanides and thiocyanates form double compounds with molybdenum in the lower valence states corresponding to the halides. These on solution in water yield similar anions containing molybdenum atoms. Characteristic of these compounds is their intense color. Their compositions are represented by substituting CN— or CNS— groups for halogens in the various complex halides, and in them, Mo atoms have a coordination number of 8. Generally, the complex molybdicyanides are somewhat more stable than the corresponding halides.

See ALLOY; FERROALLOY; IRON ALLOYS; STEEL; TRANSITION ELEMENTS. [D.H.K.]

Bibliography: R. S. Archer, J. Z. Briggs, and C. M. Loeb, Jr., *Molybdenum; Steels, Irons, Alloys*, 1948; L. Gmelin, *Handbuch der Anorganischen Chemie*, 8th ed., 1935; D. H. Killeffer and A. Linz, *Molybdenum Compounds, Their Chemistry and Technology*, 1952.

Moment of inertia

A relation between the area of a surface or the mass of a body to the position of a line. The analogous positive number quantities, moment of inertia of area and moment of inertia of mass, are involved in the analysis of problems of statics and dynamics respectively.

The moment of inertia of a figure (area or mass) about a line is the sum of the products formed by multiplying the magnitude of each element (of area or of mass) by the square of its distance from the line. The moment of inertia of a figure is the sum of moments of inertia of its parts

Moment of inertia of area. In practice, only moments of inertia of a plane area about mutually perpendicular axes (lines) in or normal to its plane are useful (Fig 1).

The moment of inertia of plane area A about the X and Y axes in its plane are respectively $I_X = \int y^2\,dA$ and $I_Y = \int x^2\,dA$. In these, x and y are the coordinate locations of area element dA.

The polar moment of inertia of a plane area is its moment of inertia about an axis normal to the plane of area. The polar moment of area A about the Z axis is $J_Z = \int r_Z^2\,dA$. As referred to a common origin of axes $J_Z = I_X + I_Y$.

Moment of inertia of area is measured in quartic

dm is the mass included in volume element dV at whose position the mass per unit volume is ρ

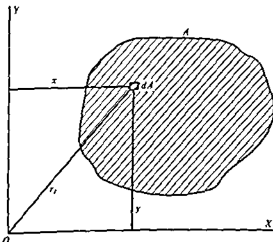


Fig. 1. Moment of inertia of an area.

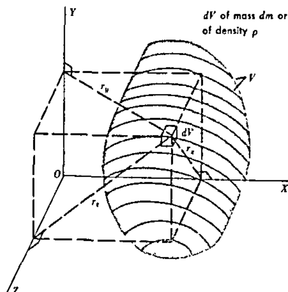


Fig. 2 Moment of inertia of a volume.

(Fig. 2). Similarly $I_Y = \int r_y^2 \rho \, dV$ and $I_Z = \int r_z^2 \rho \, dV$. Mass moment of inertia is measured in units of mass multiplying length units squared, such as g-cm².

Parallel-axis theorem. The moment of inertia of a figure about any axis is the sum of its moment of inertia about a parallel axis containing the centroid of the figure and the product formed by multiplying the magnitude of the figure (its area or mass) by the distance squared between the parallel axes; for area, $I = \bar{I} + D^2$; for mass, $I = \bar{I} + MD^2$. Accordingly, the moment of inertia about a centroidal axis is less than its moment about ...

... point are generally unequal. The moment is greatest about one and least about another perpendicular to the one. A set of three orthogonal lines consisting of these two and a line perpendicular to both are the principal axes of inertia of the figure relative to that point. If the point is *the figure's centroid*, the axes are the central principal axes of inertia. The moments of inertia about principal axes are principal moments of inertia. *See* CENTROIDS OF AREAS AND LINES; PRODUCT OF INERTIA; RADIUS OF GYRATION. [N.S.F.]

Momentum

Linear momentum is the product of the mass and the linear velocity of a body. It is defined by the equation

$$\mathbf{P} = m\mathbf{v}$$

where m is the mass and $\mathbf{v}$ is the linear velocity. Since linear momentum is the product of a scalar and a vector quantity, it is a vector and hence has both magnitude and direction.

The angular momentum of a body is defined as the product of its moment of inertia and its angular velocity, and is treated in a separate article (*see* ANGULAR MOMENTUM).

No special names are given to the units of linear momentum. The units are gram-centimeters per second, kilogram-meters per second, and slug feet per second in the centimeter-gram-second, meter-kilogram-second, and British engineering systems of units, respectively.

According to the general statement of Newton's second law, for a force **F**, a momentum **P**, and a time t,

$$\mathbf{F} = d\mathbf{P}/dt$$

Thus Newton's second law involves the time rate of change of momentum. Usually the mass of a body is constant, and the time rate of change of momentum of a body equals the product of its mass and acceleration. However, under certain conditions the mass can change, as when a rocket moves through space by consuming part of its mass as fuel. Whenever a change in mass occurs, the total time rate of change of momentum must be considered in describing the motion. Changes of momentum are important in collision processes (*see* COLLISION).

When a group of bodies is subject only to forces that members of the group exert on one another, the total momentum of the group remains constant. *See* CONSERVATION OF MOMENTUM; IMPULSE (MECHANICS). [P.W.S.]

Monazite

A phosphate (mineral) of the cerium metals, $(Ce,La,Y,Th)(PO_4)$. Ordinarily lanthanum (La) is present in about 1:1 ratio with cerium. Small amounts of the yttrium (Y) earths substitute for Ce and La. Thorium substitutes for Ce and La and generally ranges up to 10% ThO_2. A series of monazite minerals ranging up to 30% ThO_2 probably exists. Thorium-free monazite is rare. Uranium (U) in small amounts has been reported.

Monazite crystallizes in the monoclinic syste[m]. Crystals are prismatic and generally minute ... range from white throu[gh] ...

... as accesso[ry] grains and crystals in granitic and pegmatitic roc[ks] and in metamorphic gneissic rocks. In regions [of] such rock types, fluviatile and beach sands may co[n]tain commercial quantities of monazite. Monazi[te] occurs in many regions of the world, but the maj[or] production comes from placer deposits in Idah[o], South Carolina, and Florida in the United State[s] and in India, Brazil, and the Union of South Afric[a]. Recovered monazite concentrates are sources [of] thorium, thorium compounds, and cerium metal[s]. *See* CERIUM; RADIOACTIVE MINERALS; RARE EARTH ELEMENTS. [W.R.L.]

Mongolian idiocy

A type of mental deficiency in individuals, of un[...] certain cause, characterized by specific physica[l] features resembling those of Mongolian persons

The incidence among newborn is estimated at 3 in 1000 and in the general population at about 1 in 1000. The difference in incidence reflects the probability of early morbidity. For the same reason the proportion of this condition among the mentally deficient population also decreases with age, 50% during the first year, 25% at age five. Occurrence varies from 5 to 8% of mental deficients of all ages.

The characteristic physical features include the almond shaped eyes and the rounded, brachycephalic skull with flattened occipital region (cephalic index usually greater than 0.8). However, many other physical characteristics not typical of Mongolian persons are also observed. Included among these are an enlarged, fissured tongue, broad hands with stubby fingers, a short nose and depressed nasal bridge, thick, everted and cracked lips, dry rough skin, subnormal height, and infantile genitalia. All these physical signs are not present in every case and some may be observed in other varieties of mental defect. The classification of Mongoloid is applied when most of the anomalies are present.

The degree of mental defect seems to be directly related to the number and gravity of the physical signs. Despite the use of the classical generic term, idiocy, few Mongoloid cases are classified as severe. Most cases are of moderate grade. The rare cases of mild degree are educable in special classes; however, many other cases are trainable by educational definition. The personality and social behavior of this group are almost as distinctive as the physical signs are characteristic. Mongoloids are curious, observant, skillful at mimicry, and, in general, are affectionate. Aggression and hostility, which appear in many other types of mental defectives as results of frustration and inferiority feelings, are rare symptoms in this group.

Pathological study suggests nonspecific, generalized, defective brain development, thyroid dysfunction and congenital defects of the heart. Medication has little or no apparent effect on the physical condition or on degree of retardation (*see* THYROID GLAND).

Although there are reports of more than one Mongoloid in a single family, and the possibility of a genetic cause persists, it is evident that this condition is not an hereditary disease following a typical genetic pattern. The single factor which appears with consistency in study of the etiology of the condition is that the probability of a Mongoloid birth increases with maternal age. Further, the Mongoloid is late or the last in the birth order, while no significant birth order position or distribution is observed in other types of mental deficiency. Thyroid deficiency, hypopituitarism, and ovarian pathology have been observed in the mother and the probability of such an upset in endocrine balance may increase with age. However, other than the certainty that the Mongoloid is a manifestation of a defect in embryonic development, etiology is still in doubt. *See* ABNORMAL BEHAVIOR; MENTAL DEFICIENCY. [M.G.K.]

Mongoose

Any of several carnivorous animals of the family Herpestidae, somewhat similar to the weasels, but with heavier bodies, large, powerful legs, and long tails. They are native to southeastern Asia, Africa, and Spain, and have been introduced into the West Indies and Hawaii.



Golden-brown mongoose, Herpestes javanicus; length to 18 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

Mongooses are famous for their ability to catch and kill snakes, especially the dreaded cobra, although they exhibit no preference for poisonous snakes. There is no evidence that they are immune to the venom as often claimed, but survive solely through their agility.

In the West Indies, where they were introduced as a control measure against the dangerous fer-de-lance, they have become a serious enemy of poultry and other small animals. They proved equally destructive in Hawaii. *See* CARNIVORA; WEASEL. [J.D.B.]

Monhysteroidea

A superfamily of free-living nematodes with circular amphids, smooth or slightly ringed cuticle, and outstretched ovaries. Usually less than 1 millimeter in length, they occur in soil and in fresh and salt water. They are mainly herbivorous, though the presence of a buccal stylet in one marine family suggests carnivorous habits. Families within the group are distinguished by differences in the structure of the stoma and esophagus. Two genera of Monhysteridae have paired eyelike structures in the esophageal region. See NEMATODA. [H.E.W.]

Moniliaceae

A family of fungi of the order Moniliales. Mucedinaceae is the synonym for Moniliaceae. The sporophores, or fruit bodies, are usually lacking but when present they are not aggregated into fascicles, which are small bundles. The hyphae and [illegible] r [illegible] e [illegible] s [illegible] e various types of spores in the Moniliaceae are presented with their important genera.

Arthrospores. These spores are single cells formed when the hypha breaks up. *Geotrichum*

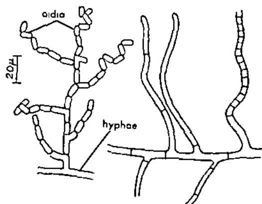


Fig. 1. *Geotrichum candidum* (*Oospora lactis*) with white mycelium. Arthrospores (*oidia*) formed by segmentation of hyphae. (*After A. de Bary, 1876*)

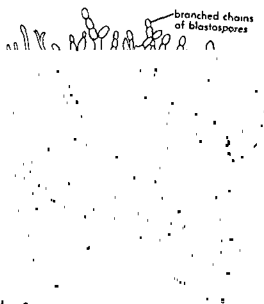


Fig. 2. *Monilia fructigena* showing branched chains of 1-celled, ellipsoid blastospores, formed acropetally by budding. (*After P. Sorauer*)

(*Oospora*) is a genus that has 1-celled, short, cylindric arthrospores with truncate ends. *G. candidum* is saprophytic and common in soil and milk (Fig. 1).

Blastospores. These are spores formed by budding either directly from a hypha or from another cell. *Monilia* is a heterogeneous genus with 1-celled catenulate, or chainlike, blastospores. Most species, such as *M. fructigena*, *M. fructicola*, and *M. laxa*, are stages of *Montilinia* (*Sclerotinia*), a genus of the Ascomycetes (Fig. 2). All of these cause brown rot, a serious disease of stone fruits particularly in the United States and a disease of minor consequence on the pome fruits. A few of the species are stages of *Neurospora* (Ascomycetes) such as *N. sitophila*.

Aleuriospores. These are terminal or lateral thick-walled nondeciduous spores made by the rounding up of a cell or cells. The genus *Trichothecium* (*Cephalothecium*) has long spore-producing bodies (sporophores) bearing clusters of 2-celled aleuriospores at the apex. *T. roseum*, a common saprophyte, sometimes causes pink rot of apples (Fig. 3). The microconidia of several genera (*Microsporum*, *Trichophyton*) of the dermatophytes are also aleuriospores.

Radula spores. These are spores on a sterigma which has no relation to the growth point of a hypha. The genus *Botrytis* has tall conidiophores, irregularly branched at the top. The ultimate branches have densely aggregated conidia, each

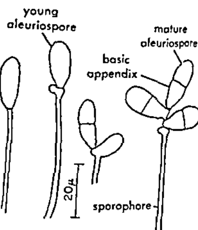


Fig 3. *Trichothecium roseum*. Sporophores of increasing age. The 2-celled pyriform conidium is considered an aleuriospore with a basic appendix (*After C. T. Ingold, 1956*)

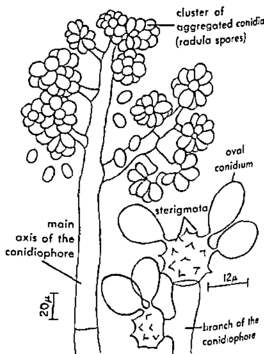


Fig. 4. *Botrytis cinerea*. Conidiophore with clusters of ovoid conidia on short sterigmata (radula spores) (*After R. E. Smith*)

at the end of a sterigma. Black sclerotia and globose or flask-shaped microconidia, called phialospores, occur. Several species are stages of *Botryotinia* (*Sclerotinia*, Ascomycetes). *B. cinerea*, a common gray mold, is a stage of *Botryotinia fuckeliana* (Fig. 4). *Beauveria* and *Botryosporium* are other genera with radula spores (Fig. 5).

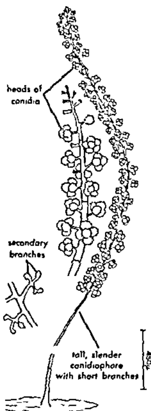


Fig. 5. *Botryosporium pulchrum.* Conidiophore tall, slender, producing numerous lateral branches, these branches producing secondary branches which bear heads of conidia on minute sterigmata. (*After A. C. J. Corda, 1839*)

Phialospores. These are conidia or spores arranged in a chain, or catenulate.

Penicillium. This is a genus with upright conidiophores, branched near the apex to form a brush-like conidial bearing apparatus. The conidia are usually 1-celled, dry, spherical, catenulate, and produced in the direction of the base (basipetally).

There are approximately 140 species recognized. A small number of these have an ascus stage (*Carpenteles*, Ascomycetes).

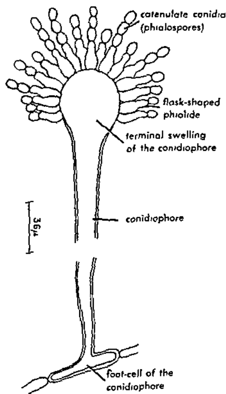


Fig. 7 *Aspergillus niveo-glaucus.* Conidiophore with radiating phialides and catenulate conidia (phialospores). (*After Thom and Raper, 1945*)

... are pathogenic and ma... bl... *P.* ... fruits. *P. notatum* and *P. chrysogenum* produce penicillin.

... and produced basipetally (Fig. 7). There are ... species recognized, and most of these are saprophytic. Some are stages of *Eurotium* (Ascomycetes)

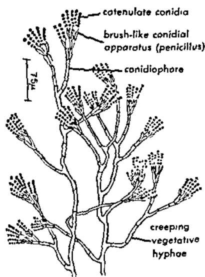


Fig. 6. *Penicillium expansum.* Conidiophores arising from the creeping hyphae. (*After O. Brefeld, 1874*)

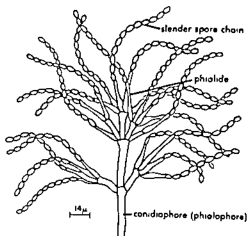


Fig. 8. *Spicaria (Paecilomyces) elegans* The genus *Spicaria* is much like *Penicillium*, but the spore-bearing apparatus is less compact, the phialides more spreading and the spore-chains longer (After H. S. Brown and G. Smith, 1957)

A. (*Sterigmatocystis*) *niger* causes the common black mold of cotton bolls, certain fruits such as figs and dates, and vegetables; *A. fumigatus* causes aspergillosis in man and birds.

Spicaria. This is a poorly defined genus having many features in common with *Penicillium* and *Paecilomyces*. However, the phialides spread to a greater degree. *S. elegans* is a common soil inhabitant (Fig. 8).

Cephalosporium. This genus has conidiophores which are in reality phialides, arising as branches from all parts of the creeping hyphae. The conidia are 1-celled ...

... genus with upright conidiophores, bearing lateral tapering phialides arranged in whorls. The conidia are 1-celled. Species are mostly parasites on vascular plants. *V. albo-atrum* causes wilt diseases of many plants. *See* FUNGI; MONILIALES.

[N.F.B.]

Moniliales

An order of fungi of the Fungi Imperfecti containing many plant pathogens. The Moniliales are divided into four families, the Moniliaceae (Mucidinaceae), Dematiaceae, Stilbellaceae (Stilb...

... spores are always formed free on the surface of the material on which the organism is living and never occur in either pycnidia or acervuli. The pycnidium is a globose or flasklike fruit body containing pycnidiospores. The acervulus is a cushionlike mass of hyphae having conidiophores and conidia. There are two main types of asexual spores, thallospores and true conidia. Thallospores are formed by the transformation of preexisting elements of the thallus or vegetative hyphae (Fig. 1). There are three kinds of thallospores: (1) arthrospores, or oidia, formed by the breaking up of a hypha into cells, as in *Geotrichum*; (2) blastospores, or sprout cells, arising as buds from preexisting cells, as in true yeasts, for example, *Monilia* and *Candida*; and (3) aleurospores, thick-walled nondeciduous spores, always on the end of the hyphae which bear them and resembling conidia in situation, color, and form

The true conidia arise on the thallus as newly formed elements and are always deciduous, that is, they do not remain on the hyphae (Fig 2). There are two kinds of true conidia: (1) radula spores borne on small projections, or sterigmata, from a

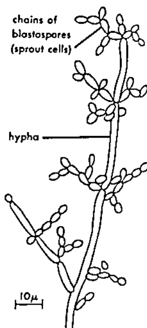


Fig. 1. Thallospores. *Candida albicans*. Hypha with blastospores. (After Diddens and Lodder, 1942)

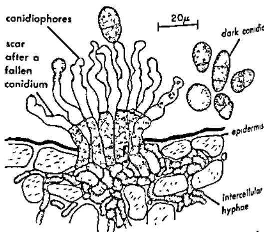


Fig. 2. True conidia. *Polythrincium trifolii*. Cluster of irregularly bent conidiophores producing dark, 2-celled conidia. (After F. A. Wolf, 1935)

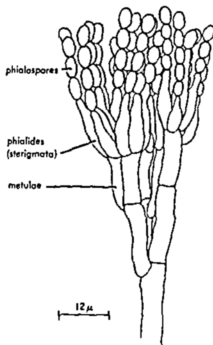


Fig 3. Phialospores. *Penicillium expansum* Conidiophore with flask-shaped phialides producing 1-celled, catenulate conidia (phialospores). (*After Raper and Thom, 1949*)

cell, as in *Botrytis* and (2) phialospores borne on phialides as in *Penicillium* (Fig. 3). The phialides are terminal, often flask-shaped cells from the interior of which a basipetal succession of conidia develops [N.F.B.]

Monitoring (ionizing radiation)

The use of meters and special techniques to determine the absorbed dose of ionizing radiation received by individuals; also, the use of meters and other devices to determine the rate and intensity of radiation in various areas. Radiation monitoring is accompanied by many other health physics services and functions. For example, if a health physicist is assigned to a radiation survey or monitoring operation, he will not only measure the dose or dose rate but also will put into effect measures to minimize the exposure. For example, he may specify required shielding, indicate necessary decontamination procedures, or advise the use of appropriate remote control equipment, protective clothing, or glove boxes. Monitoring frequently is divided into three categories: personnel monitoring, building surveys, and area monitoring.

Personnel monitoring. This involves those operations directly associated with the measurement and recording of the absorbed dose received by the individual and all health physics services and functions designed to minimize the exposure to the individual Personnel monitoring includes the issuing of dosimeters and film badges; the reading, maintenance, and calibration of these devices; the keeping of exposure records; and personal contacts with individuals to determine the causes of exposure and to recommend procedures to limit the recurrence of exposures. Personnel monitoring includes the use of such things as hand and foot counters to measure hand and foot contamination, friskers to measure clothing contamination, and probe counters to measure wound contamination. Many personnel monitoring programs require the use of decontamination laundry facilities to decontaminate clothing, the use of total body counters to estimate the buildup of radioactive contamination inside the body, and the operation of a chemical analysis laboratory to measure the level of radioactive material in the blood, urine, and feces. *See* DECONTAMINATION (RADIOACTIVE CONTAMINANTS); DOSIMETER; FILM BADGE.

Building surveys. These are made with many types of survey or monitoring instruments which, for the most part, can be grouped into three classes: (1) Geiger-Müller counters, (2) scintillation counters, and (3) ionization chambers. This equipment is used to measure the dose rate and accumulated dose in various work areas and to estimate the surface contamination on floors, walls, furniture, and equipment Many refinements have been made in instruments and techniques to measure separately the doses from x-rays, α-rays, β-rays, γ-rays, fast neutrons, and other types of ionizing radiation It is not sufficient to measure just the total dose of ionizing radiation, for each component of the mixture may have a different relative biological effectiveness (RBE). One of the hazards of greatest concern in many types of work with radioactive materials is the inhalation of airborne dusts, fumes, and gases As a consequence, various types of equipment, such as air filters, precipita-

tors are operated continuously in various areas, making a record of the dose rate and air contamination, or both, on a graphic recorder. A warning alarm is sounded if the radiation level or the air contamination exceeds certain prescribed safe values.

Area monitoring. This is concerned with the measurement of the buildup and spread of radioactive contamination in the air, water, and soil close to work areas. Many of the instruments used for building surveys are used in area monitoring. In addition, fall-out trays are used to collect the dust that settles to the ground, rain samplers are used to collect and measure the radioactive contamination in the rain, and various types of probes are used to measure the level of ground-water contamination in sampling wells and to measure the contamination in river waters and sediment. Area monitoring and ecological studies begin in an area be-

tion and the accumulation of radioactive materials in the rivers and soil and in plants and animals may be determined and continuously monitored. *See* GEIGER-MÜLLER COUNTER; *IONIZATION CHAMBER*; PARTICLE DETECTOR; RADIOACTIVE FALLOUT; SCINTILLATION COUNTER. [K.Z.M.]

Monkey

Any of a large number of mammals of the order Primates, divided into two major groups, the New World monkeys, superfamily Ceboidea; and the Old World monkeys, superfamily Cercopithecoidea.

The New World monkeys are distinguished by the absence of cheek pouches; the separation of the nostrils by a broad, flattened area; and in some, by a prehensile tail. The New World monkeys are abundant in Central and South America, varying greatly in details and size. The little marmosets, genus *Hapale*, are scarcely larger than squirrels and make excellent pets. In contrast are the moderately large howlers, genus *Alouetta*, whose loud and varied sounds are familiar in the American jungle.





(a) Common macaque, Macacus cynomolgus. (b) Priamus monkey, Semnopithecus nestor; length to 16 in. (From P. M. Duncan, ed., Cassell's Natural History, Cassell)

Old World monkeys never have a prehensile tail, their nostrils are set close together and point downward; and many have internal cheek pouches. They also have bare, calloused areas of skin on their buttocks. Among the Old World monkeys, the baboons and the rhesus monkeys of the genus *Macacus* are best known. The latter are of great importance in medical research because of their availability and their physiological similarity to man. *See* BABOON; PRIMATES. [J.D.]

Monocotyledoneae

A subclass of the plant class Angiospermae, which comprises approximately 50,000 species. They are herbaceous plants except for the palms and a few

bium and are enclosed by cells which collectively make up a bundle sheath). The leaves are usually parallel-veined from base to tip, mostly with the margins entire (without teeth or divisions). The flower parts are basically in 3s or multiples of 3. The embryos have a single cotyledon (*see* PLANT ANATOMY). Monocotyledonous plants—wheat, rice, corn, and other grasses—are the most important food plants in the world.

It is thought that both the Monocotyledoneae and the higher Dicotyledoneae developed from some primitive genetic group of the dicotyledons, perhaps the order Ranales. The Monocotyledoneae is frequently divided into 10 orders: Pandanales, Helobiales, Graminales, Palmales, Synanthales, Arales, Farinales, Liliales, Scitaminales, and Orchidales. See separate articles describing each of these orders; *see also* ANGIOSPERMAE; PLANT KINGDOM. [P.D.S.]

Bibliography: *See* EMBRYOPHYTA.

Monogenea

A subclass of the Trematoda which are ectoparasites of the gills, skin, and orifices of fishes and, less frequently, of the esophageal tracts and bladders of amphibians and turtles. They have conspicuous anterior and posterior holdfasts, the latter usually armed. The terminal genitalia are frequently sclerotized. The group is characterized by sexual reproduction, direct development, and a single host in the life cycle. Egg capsules have terminal filaments.

Taxonomy. The distinctive shapes and sclerotized holdfasts facilitate characterization and arrangement. The most widely used classification employs two orders, the Monophisthocotylea, in which the posthaptor is without discrete multiple suckers or clamps, and the Polyopisthocotylea with suckers or clamps on the posthaptor.

Morphology. Body shapes of the various genera are distinctive, sometimes bizarre as in *Vallisa* which is c-shaped. Paired external suckers or buc-

cal cavity suckers and adhesive glands occur anteriorly. The posterior holdfast is either solid and *armed with central anchors and marginal hooks* (Gyrodactyloidea), sucker-shaped with anchors and hooks (Capsaloidea), or solid and bearing suckers or clamps (Polyopisthocotylea). The pharynx is muscular and protrusible, and the esophagus is short, often ramified. The gut is commonly bifid, often much ramified. Multiple testes and numerous vitelline glands are usually present. A genitointestinal canal occurs in many, and the penes are usually *armed*.

Life history. Monogenea usually have direct development involving simple metamorphosis from the ciliated larval stage to the nonciliated juvenile. Juvenile anchors and hooks may be retained or replaced by adult suckers or clamps. Cross-fertilization or, perhaps less frequently, self-fertilization of

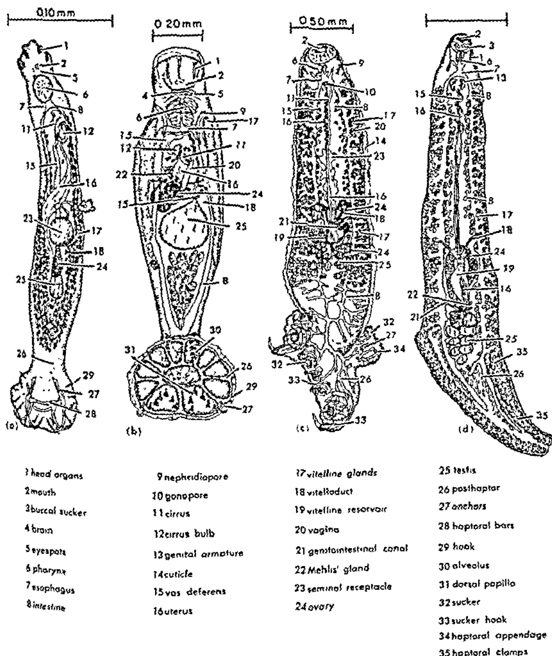


Fig. 1. Representative monogeneids. (a) Superfamily Gyrodactyloidea, *Pseudohaliotrema carbunculus* Hargis from *Lagodon rhomboides*, the pinfish, dorsal view. (b) Superfamily Capsaloidea, *Heterocotyle aetobatis* Hargis from *Aetobatis narinari*, the spotted eagle ray, ventral view. (c) Superfamily Polystomatoidea, *Heteronchocotyle leucas* Hargis from *Carcharhinus leucas*, the bull shark, ventral view. (d) Superfamily Diclidophoroidea, *Heteraxine xanthophilis* Hargis from *Leiostomus xanthurus*, the spot, ventral view

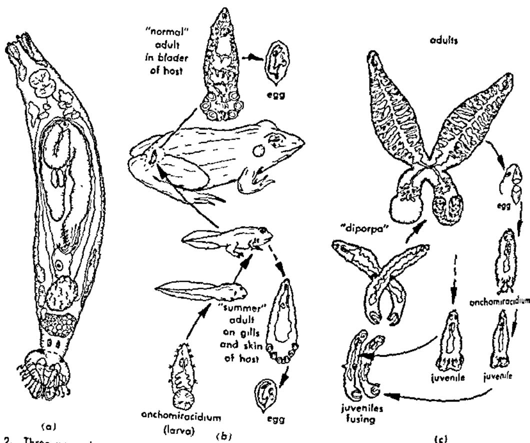


Fig. 2. Three unusual monogeneid life cycles. (a) *Gyrodactylus funduli* Hargis from *Fundulus similis*, the longnose killifish, dorsal view. In the uterus of the viviparous parent, thought by some to be polyembryony, is an embryo within an embryo. (b) *Polystoma integerrimum* (Frolich) from the frog bladder. The larva on tadpole gills grows as the tadpole develops. If, when the tadpole metamorphoses, the juvenile fluke does not enter the host's esophagus whence it can travel to the bladder, it matures neotenically into a "summer" adult and reproduces the first year. If, however, it does reach the frog's bladder, the fluke matures normally and reproduces in the third year. (c) *Diplozoon* sp. The larvae undergo simple metamorphosis into nonciliated juveniles which attach to each other and become fused "diporpa larvae." Fusion of reproductive organs becomes complete as the flukes mature and cross fertilization occurs as the occasion demands.

hermaphroditic individuals resulting in egg capsules which hatch on the host or in its environment is general; however, interesting exceptions exist.

Physiology. Until recently, monogeneids were thought to feed on mucus but there is good evidence that some are blood feeders. Most live in well-aerated microhabitats and probably require more oxygen than entoparasites. The male and female organs of some species mature at different times but the reproductive significance is uncertain.

Ecology. Monogeneids are capable of defi... togenetic ...

rela

T

... well accommodated in their hosts, monogeneids may cause intense discomfort and wasting and pave the way for invasion by bacteria, myceliated fungi, and protozoans. Little is known of their effects on healthy, natural populations. Heavy infections cause serious mortality in hatcheries, aquaria, and culture ponds and sometimes occur in nature. See ASPIDOCASTREA; DIGENEA; TREMATODA. [W.J.HA]

Bibliography: N. G. Sproston, A synopsis of the monogenetic trematodes, *Zool. Soc. (London)*, 25(4):185–600, 1946.

Monogononta

An order of the class Rotifera which contains the majority of species in the class. The organisms of this order are characterized by the presence of a single gonad in both males and females. There is a striking degree of sexual dimorphism, with the males being small and degenerate. The order is made up of three suborders: Ploima, Flosculariacea and Collothecacea.

In suborder Ploima, there is an exceptional diversity of form, varying from soft-bodied wormlike rotifers to species with variously ornamented, lori-

cate shells. Most of the free-swimming benthonic and pelagic rotifers belong to this suborder. Locomotion is by the ciliated corona. No fewer than five kinds of trophi occur: virgate (notommatids), cardate (*Lindia*), forcipate (dicranophorids), malleate (brachionids) and incudate (asplanchnids). Species range in size from 0.05 mm, in *Colurella*, to 1.60 mm in *Asplanchna*.

The suborder Flosculariacea contains the spectacular sessile rotifers formerly known as melicertaceans of the family Flosculariidae as well as a number of equally notable free-swimming forms in-

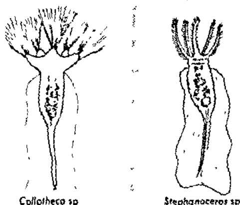


Fig 3 Collothecacea.

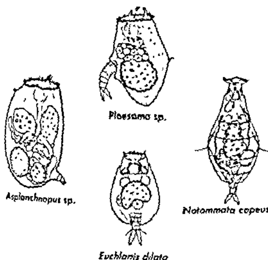
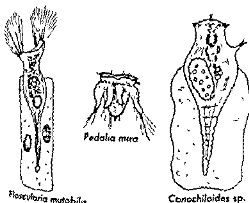


Fig 1. Ploima.



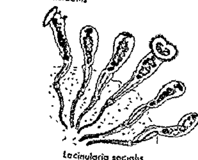


Fig 2 Flosculariacea.

cluded in the family Testudinellidae. Among the latter are the globular *Trochosphaera*, hailed as a connecting link between rotifers and annelids, and the appendage-bearing *Pedalia*, once considered as transitional to crustacea. All members of this suborder have a malleoramate mastax. The flosculariids often have free-swimming young, and a few kinds are free-swimming as adults Some of the sessile forms are encased in intricately constructed tubes Some species occur in spherical or ellipsoidal colonies containing many individuals. Colonies may be over 2.0 mm in diameter, but individual flosculariians are seldom over 1.0 mm in length.

The suborder Collothecacea contains but a single family, the Collothecidae, made up of five genera. Most species of Collothecidae are sessile, and many are encased in gelatinous tubes. The largest genus, *Collotheca*, with an expanded funnel-shaped anterior end, has a wide range of coronal shapes among its 50 or more species In its simplest form, the corona of *Collotheca* is without lobes, but in some species the corona has 5–7 lobes, which may be long and fingerlike. Three genera lack a corona. All members of this suborder have an uncinate mastax. *Collotheca hoodi*, the largest rotifer known, with the exception of Seisonacea, belongs to this group and reaches length up to 2.5 mm. *See* ROTIFERA. [E.H.A.]

Monoplacophora

The most recent major group of mollusks to be recognized. They are important in providing new evidence on the structure of the primitive molluscan stock and on the interrelations of the five currently recognized classes of mollusks. Many workers consider them to be a sixth class of mollusks; a few consider them to be an order allied to the Polyplacophora (Amphineura). *See* MOLLUSCA.

Geologic history. The Monoplacophora were first recognized in the fossil record. Simple cap-shaped fossil shells resembling those of Recent rock-clinging gastropods have been known for more than 125 years. For over 75 years it has been observed that lower Paleozoic shells of this shape have discrete

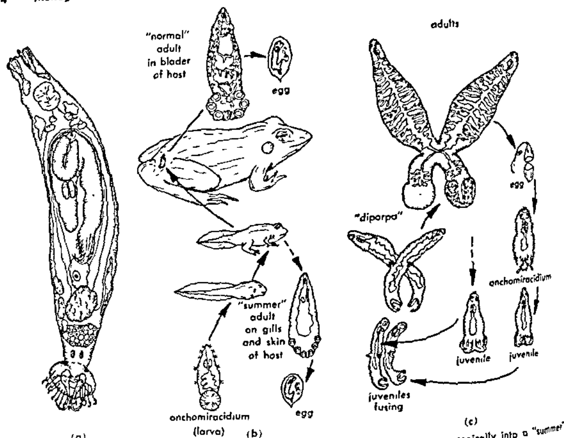


Fig. 2. Three unusual monogeneid life cycles. (a) *Gyrodactylus funduli* Hargis from *Fundulus similis*, the longnose killifish, dorsal view. In the uterus of the viviparous parent, thought by some to be polyembryony, is an embryo within an embryo. (b) *Polystoma integerrimum* (Frolich) from the frog bladder. The larva on tadpole gills grows as the tadpole develops. If, when the tadpole metamorphoses, the juvenile fluke does not enter the host's esophagus whence it can travel to the bladder, it matures neotenically into a "summer" adult and reproduces the first year. If, however, it does reach the frog's bladder, the fluke matures normally and reproduces in the third year. (c) *Diplozoon* sp. The larvae undergo simple metamorphosis into non-ciliated juveniles which attach to each other and become fused "diporpa larvae." Fusion of reproductive organs becomes complete as the flukes mature and cross fertilization occurs as the occasion demands.

hermaphroditic individuals resulting in egg capsules which hatch on the host or in its environment is general; however, interesting exceptions exist.

Physiology. Until recently, monogeneids were thought to feed on mucus but there is good evidence that some are blood feeders. Most live in well-aerated microhabitats and probably require more oxygen than entoparasites. The male and female organs of some species mature at different times but the reproductive significance is uncertain.

Ecology. Monogeneids are capable of delicate ontogenetic responses to the microhabitat; for example, the direction of body asymmetry often depends upon which side of the host the worm inhabits. The direct life history seems to favor a higher degree of host specificity than in other trematodes. Often a parasite occurs on a single host species, and closely related fishes bear closely related parasites.

Though usually well accommodated in their hosts, monogeneids may cause intense discomfort and wasting and pave the way for invasion by bacteria, myceliated fungi, and protozoans. Little i known of their effects on healthy, natural popula tions. Heavy infections cause serious mortality in hatcheries, aquaria, and culture ponds and some times occur in nature. *See* ASPIDOGASTREA; DI GENEA; TREMATODA. [W.J.H.]

Bibliography: N. G. Sproston, A synopsis of the monogenetic trematodes, *Zool. Soc. (London)*, 25(4):185–600, 1946.

Monogononta

An order of the class Rotifera which contains the majority of species in the class. The organisms of this order are characterized by the presence of a single gonad in both males and females. There is a striking degree of sexual dimorphism, with the males being small and degenerate. The order is made up of three suborders: Ploima, Floscularia cea and Collothecacea.

In suborder Ploima, there is an exceptional diver sity of form, varying from soft-bodied wormlike rotifers to species with variously ornamented, lori-

broad, short animal with a poorly developed body cavity (coelom) and incomplete pairing of major organs. The mollusks may have developed by losing segmentation, the annelids by concentrating on segmentation of the body, and the arthropods by concentrating on segmentation of appendages. It will take years for the new ideas supplied by *Neopilina* to be evaluated. There is no question, however, that its discovery is a significant find for zoology and a spectacular proof of an interpretation of paleontological information. [E.L.Y.]

Monopulse radar

A radar that obtains a complete measurement of the target's angular position from a single echo pulse. Together with the range measurement performed with the same pulse, the target position in three dimensions is determined completely. Usually a train of echo pulses is then employed to make a large number of repeated measurements and produce a refined estimate, but this is not intrinsically necessary.

The antenna receiving characteristics are especially pertinent to monopulse performance. It is usually convenient, but not necessary, to use the same antenna to illuminate the target. The monopulse operation is implemented by means of two pairs of feed points appropriately located at the antenna. The two feeds for determining azimuth information are located in the same horizontal plane on either side of the beam axis. The main lobe due to one is directed slightly to the left of the beam axis, and the main lobe belonging to the other is tilted slightly to the right. A target located exactly on the beam axis produces the same signal in both feeds, but a target located to one side of the beam axis produces a stronger signal in one feed than the other. The difference between the signals received by the two feeds indicates the azimuth separation between the beam axis and the line of sight to the target. The sum of the signals received by the two feeds indicates the gross signal strength and is used as a normalizing factor. See ANTENNA (AERIAL).

A similar arrangement is used to measure elevation angle, with the two feeds located in the same vertical plane above and below the beam axis. There is nothing to prevent simultaneous measurement of azimuth and elevation.

The same antenna can also be used for transmission by connecting the four feed points in parallel. The radiation pattern thereby produced is almost the same as the pattern produced by a single feed, except that the beamwidth is slightly smaller, the gain slightly higher, and the lobe pattern is symmetrically positioned with respect to the beam axis instead of being tilted.

The sum and difference signals can be formed either at the carrier frequency or after conversion to a lower frequency. It is of some advantage to obtain the sum and difference results immediately at the carrier frequency to minimize the possibility of subsequent circuits introducing errors. The operations can be performed conveniently at microwave frequencies by the use of wave-guide hybrid mixers. Only the sum and difference signals, in both azimuth and elevation, would then be conveyed beyond this point in the receiver

An important advantage of a monopulse tracking radar over one employing conical scan is that the instantaneous angular measurements are not subject to errors caused by target scintillation. In a conical-scan system the angular information is derived from the phase and amplitude of the modulation envelope on the echo pulse train. The conical nutation frequency usually lies in the range between 20 and 40 cps, and the scintillation spectrum of most targets possesses nonnegligible components within this band. The scintillation produces random modulation at the nutation frequency which can not subsequently be differentiated from the signal, leading to tracking errors A monopulse tracking radar is not susceptible to this difficulty because each pulse provides an angular measurement without regard to the rest of the pulse train. There is no opportunity for radar cross-section fluctuations to affect the measurement.

An additional advantage of monopulse tracking as compared to conical scan is that no mechanical action is required in monopulse. In a precise and highly maneuverable tracking radar it is helpful not to have to accommodate a whirling scanner For discussion of conical scan *see* RADAR. [R.L.B.]

Bibliography: D. R. Rhodes, *Introduction to Monopulse*, 1959.

Monorail

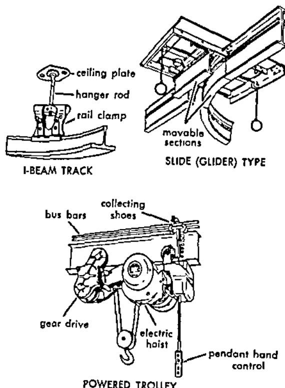
A distinctive type of materials handling machine that provides an overhead, normally horizontal, fixed path of travel in the form of a trackage system and individually propelled hand or powered trolleys which carry their loads suspended freely and moves them with intermittent motion. This last characteristic differentiates them from overhead conveyors Because they operate over fixed paths rather than over limited areas, they differ from overhead-traveling cranes, nor should they be confused with such overhead conveyors as cableways.

Relatively simple but adequately efficient monorail systems for specialized applications have flat steel bars or galvanized pipes for trackage. Monorails in packing plants consist of 2½ in. × ⅜ in. or 2½ in. × ½ in. flat stock with overhead supports, hangers, and switches. Single- or two-wheeled trolleys carry hooks for carcass meat, fowl, and large cuts; special carriers carry sausages and hams.

Pipe rails are used in systems for handling garments on hangers or wheel-equipped trolleys. Switches, crossovers and other components make setups flexible.

Both of the above systems can be arranged to run onto freight elevators, along loading platforms and directly into carriers.

Of primary importance in a conventional monorail system are the rails. They are connected with



Typical parts of overhead monorail.

butt or lap joints. Clamps and brackets suspend the trackage from ceiling or walls (see illustration). Because monorail systems do not operate with continuous motion, the tracks need not be arranged in self-closing lines. Spurs can be run into paint booths, cooling rooms, and similar work areas. These paths are selected by switches, such as the tongue or the slide (glider) varieties. Specially constructed sections permit 90° changes in trolley travel; turntables permit articles to be turned through 360°. Where monorails pass through fire doorways, lift-out sections permit the door to close in the event of fire. Lift and drop sections shift the flow of traffic from one line to another at a different level, thus eliminating the need for inclined and declined tracks.

Wheels mounted in trolleys ride on the flanges of the track in most varieties. Two-wheeled trolleys, connected by bars, combine to make up carriers. The number of wheels depends upon the desired load-carrying capacity. In many instances hand or powered hoists are suspended from carriers. To provide current for powered hoists, special electrification equipment such as bus bars and collecting shoes are used. If powered carriers are required, the drive may be secured in one of several ways. Examples are the ... which cor... trolley wh... by an elec... motor. Trolleys of the first type operate without slippage and make possible the introduction of sections with slight inclines and declines. Special carriers are used with monorail systems for handling specific products such as batches in bakeries, and the movement of clothes from one process to another in laundries. Scale sections can be included in the trackage line to weigh such products as batches, textile beams, or rolls of paper without delays in traffic flow. Many below-the-hook devices can be used with monorails (*see* HOISTING MACHINES). Monorail elements are utilized in the construction of overhead-traveling cranes if distribution over an area rather than along a fixed path is required. Such cranes are not capable of handling the heavy loads carried by cranes built with structural girders, but they are adequate for many industrial applications. *See* MATERIALS HANDLING MACHINES; *see also* BULK-HANDLING MACHINES. [D O R.]

Bibliography: Monorail Manufacturers Association, *Specifications for Overhead Track Systems*, 1955.

Monosaccharide

A class of simple sugars containing a chain of 3–10 carbon atoms in the molecule, known as polyhydroxy aldehydes (aldoses) or ketones (ketoses). They are very soluble in water, sparingly soluble in ethanol, and insoluble in ether. The number of monosaccharides known is approximately 70, of which about 20 occur in nature, and the remainder are synthetic. The existence of such a large number of compounds is due to the presence of asymmetric carbon atoms in the molecules. Aldohexoses, for example, which include the important sugar glucose, contain no less than four asymmetric atoms, each of which may be present in either D or L configuration. The number of stereoisomers rapidly increases with each additional asymmetric carbon atom. *See* CARBOHYDRATE; KETONE; OPTICAL ACTIVITY; STEREOCHEMISTRY.

A list of the best known monosaccharides is given below:

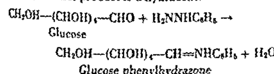
Trioses: $CH_2OH\ CHOH\ CHO$, glycerose (glyceric aldehyde)
$CH_2OH{\cdot}CO\ CH_2OH$, dihydroxy acetone
Tetroses: $CH_2OH\ (CHOH)_2\ CHO$, erythrose
$CH_2OH\ CHOH\ CO\ CHO$, erythrulose
Pentoses: $CH_2OH\ (CHOH)_3\ CHO$, xylose, arabinose, ribose
$CH_2OH\ (CHOH)_2\ CO\ CH_2OH$, xylulose, ribulose
Methyl pentoses (6-deoxyhexoses):
$CH_3(CHOH)_4CHO$, rhamnose, fucose
Hexoses: $CH_2OH\ (CHOH)_4\ CHO$, glucose, mannose, galactose
$CH_2OH{\cdot}(CHOH)_3\ CO\ CHOH$, fructose, sorbose
Heptoses: $CH_2OH{\cdot}(CHOH)_5\ CHO$, glucoheptose, galamannoheptose
$CH_2OH\ (CHOH)_4\ CO\ CH_2OH$, sedoheptulose, mannoheptulose

Aldose monosaccharides having 8, 9, and 10 carbon atoms in their chains have been synthesized

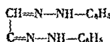
Reactions. As polyhydroxy aldehydes or ketones, the monosaccharides have numerous reactions.

Reduction and oxidation. On reduction, the aldoses take up two atoms of hydrogen and are converted to the corresponding sugar alcohols. A pentitol or pentahydric alcohol is obtained from a pentose, and a hexitol or hexahydric alcohol is obtained from a hexose. On oxidation, the monosaccharides yield carboxylic acids. Mild oxidation converts aldoses first into the corresponding monocarboxylic acids with the same number of carbon atoms; thus, aldopentoses are transformed into pentonic acids, $CH_2OH\cdot(CHOH)_3\cdot COOH$, and aldohexoses into hexonic acids, $CH_2OH\cdot(CHOH)_4\cdot COOH$. With stronger oxidizing agents, the process may proceed further, and hexoses, for example, may be oxidized to the corresponding isomeric, saccharic, or tetrahydroxyadipic acids, $COOH\cdot(CHOH)_4\cdot COOH$. The ketoses, on oxidation, yield acids containing a smaller number of carbon atoms. The reducing properties of the monosaccharides are shown by their behavior with ammoniacal silver nitrate solution, from which metallic silver is precipitated, and particularly with Fehling's solution from which, on warming, a brick red precipitate of cuprous oxide is formed. This behavior is characteristic of aldoses as well as ketoses.

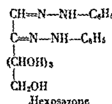
Phenylhydrazine. Emil Fischer's 1884 introduction of phenylhydrazine, $C_6H_5NHNH_2$, as a reagent in sugar chemistry has proved of the greatest value in the separation and identification of the various monosaccharides. When 1 mole of phenlhydrazine reacts with 1 mole of an aldose or ketose sugar, the first product is a hydrazone.


$$CH_2OH-(CHOH)_4-CHO + H_2NNHC_6H_5 \rightarrow$$
Glucose
$$CH_2OH-(CHOH)_4-CH{=}NHC_6H_5 + H_2O$$
Glucose phenylhydrazone

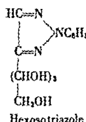
On warming with excess of phenylhydrazine, the hydrazone first formed is oxidized in such a way that the CHOH group adjacent to the original aldehydic or ketonic group is converted into a CO group. The latter then combines with another mole of phenylhydrazine to give a dihydrazone containing the group



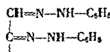
These compounds are termed osazones.


Hexosazone

The osazones are colored compounds which are difficult to purify. For this reason, their melting points and specific rotations cannot be relied upon. However, since the osazones produced by various sugars possess characteristic crystalline forms, they are frequently used for cursory identification of the parent sugar by examining the crystals under the microscope. The osotriazoles obtained by oxidation of the osazones with copper sulfate are to be preferred, because they are colorless crystalline compounds and have more definite melting points and higher specific rotations than the osazones.

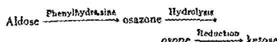

Hexosotriazole

Osazones, like all hydrazones, are hydrolyzed when heated with hydrochloric acid, resulting in regeneration of phenylhydrazine. The original sugar, however, is not recovered, because in the process of regeneration of the sugar, the group

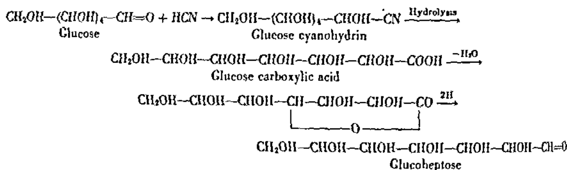


is converted into the group —CO—CHO. The highly reactive compound so formed is an oxidation product of the original sugar and is termed osone. In the example quoted above, the osazone of glucose yields glucosone, $CH_2OH\cdot(CHOH)_3\cdot CO\cdot CHO$.

The phenylhydrazine residues from sugar osazones may also be removed with a competing aldehyde, such as benzaldehyde or acetaldehyde, resulting in osone formation. On mild reduction of this compound with zinc dust and dilute acetic acid, the aldehydic group alone is attacked and converted into an alcoholic group, the keto group remaining unchanged. In the case of glucosazone, the sugar finally obtained is fructose, $CH_2OH\cdot(CHOH)_3\cdot CO\cdot CH_2OH$, in place of the glucose used as starting material. These reactions may be used as a general method of transforming an aldose into a ketose, according to the scheme


$$\text{Aldose} \xrightarrow{\text{Phenylhydrazine}} \text{osazone} \xrightarrow{\text{Hydrolysis}} \textit{osone} \xrightarrow{\text{Reduction}} \textit{ketose}$$

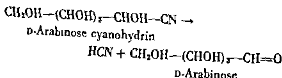
Cyanohydrin synthesis. Monosaccharides, such as aldehydes and ketones, react with hydrogen cyanide to form cyanohydrins. By the use of this reaction, which is due to H. Kiliani, the synthesis of a higher from a lower aldose can be effected. The cyanohydrins are first hydrolyzed to hydroxy acids, which are readily converted into lactones. The latter are then reduced to aldoses by means of sodium amalgam. Thus, glucose, under these conditions, results in a new seven-carbon sugar, glucoheptose;



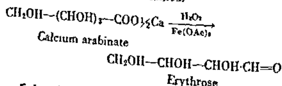
Similarly, using the glucoheptose and continuing with the process of cyanohydrin synthesis, an octose can be obtained. The synthesis has been carried as far as glucodecose.

Since cyanohydrin synthesis introduces a new asymmetric carbon atom, two products are obtained from a single monosaccharide. As an example, both L-gluconic and L-mannonic acids are ...

... for effecting degradation of an aldose to one of lower carbon content. Thus, when the aldoxime of D-glucose, CH_2OH—$(CHOH)_4$—$CH{=}NOH$, is heated with acetic anhydride, it is converted into the acetyl derivative of the *nitrile*, $CH_2OAc\cdot(CHOAc)_4\cdot CN$. On treatment with ammoniacal silver nitrate this ... the *cyanide* ... *dopento* ... al-

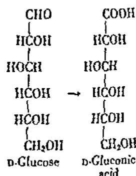


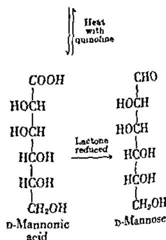
It is possible also to reduce the number of carbon atoms in an aldose carbon chain by O. Ruff's method, in which the calcium salt of an aldonic acid is oxidized with hydrogen peroxide in the presence of ferric acetate. Thus, from calcium arabinate, erythrose is obtained.



Epimerization. When two sugars or their derivatives, such as sugar acids, differ only in the configuration of the substituents on the carbon atom adjacent to the reducing group, they are called epimers. The aldonic acids are noteworthy for the ease with which they undergo epimerization or partial inversion of the asymmetry at carbon atom 2 upon heating with a weak base, such as pyridine or quinoline, to produce a mixture of the two epimers.

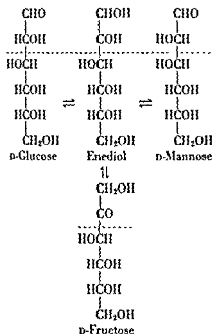
An aldose is first oxidized to the corresponding monocarboxylic acid, which is then heated with aqueous quinoline or pyridine, to yield a mixture of the epimeric aldohexonic acids. The latter may be separated by fractional crystallization of their lactones. Reduction of the lactone yields the corresponding aldose. This process may be illustrated by the transformation of D-glucose to D-mannose:





Enolization. Treatment of an aldose sugar with dilute alkali results in a mixture of an epimeric pair and a 2-ketohexose. For example, when either D-glucose, D-mannose, or D-fructose is used, a mixture of these three sugars is obtained. The reaction, known as the Lobry de Bruyn-Ekenstein transformation, which is of general application, is due to the production of enolic forms in the presence of hydroxyl ions, followed by a rearrangement as shown on the facing page.

The transformation serves to show the close relationship between glucose, fructose, and mannose;

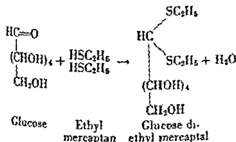


the structural representations of these three sugars are identical below the dotted line of the formula

In a similar manner, the D-galactose series yields a mixture containing D-galactose, its epimer D-talose, and the ketose D-tagatose. When either D-xylose, D-lyxose, or D-xylulose is used, a mixture of the three sugars is obtained.

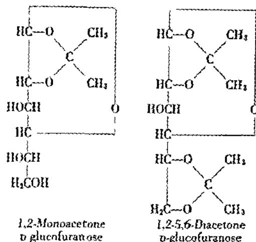
Derivatives. The derivatives of the monosaccharides are discussed in this section.

Sugar mercaptals. Reducing sugars, except ketoses, react with mercaptans in the presence of concentrated hydrochloric acid to form mercaptals. The reaction with glucose and ethyl mercaptan is given as an illustration:

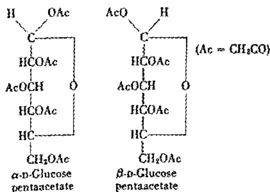


The mercaptals of the sugars are well-defined crystalline compounds. They are of special interest because they are open-chain compounds and have been found useful for the preparation of other derivatives that have this structure.

Acetone sugars. The reducing sugars yield condensation products with aldehydes and ketones. Those with acetone have played an important role in solving problems of sugar structure. The acetone glucoses are obtained by treating glucose in acetone with a condensing reagent, such as zinc chloride or sulfuric acid. The following formulas have been assigned to mono- and diacetone D-glucose, both of which are furanose derivatives:

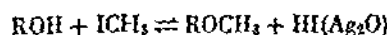


Acetylated sugars. Acetylation of all free hydroxyls resulting in the formation of an ester may be accomplished by heating the monosaccharide with acetic anhydride in the presence of a catalyst such as anhydrous sodium acetate or zinc chloride, or by treating the sugar with acetic anhydride in a pyridine solution. When the hydroxyl at carbon 1 of an aldose is acetylated, two isomers (α,β) may be produced The directive influence of the catalyst used in the reaction is important: glucose acetylated with acetic anhydride in the presence of zinc chloride gives the α-pentaacetate, whereas with a sodium acetate catalyst the β form is produced Furthermore, the β form, on heating with zinc chloride, is converted into the α isomer.

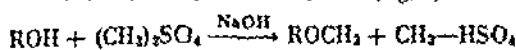


The free sugar may be recovered by deacetylation with dilute sodium hydroxide, or by catalytic deacetylation with barium or sodium methoxide.

Methylated sugars. Treatment of sugars with methyl iodide and silver oxide or with dimethyl sulfate and sodium hydroxide results in the formation of methylated derivatives (ethers) according to the following reactions:



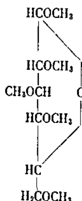
$$ROH + ICH_3 \rightleftharpoons ROCH_3 + HI(Ag_2O)$$



$$ROH + (CH_3)_2SO_4 \xrightarrow{NaOH} ROCH_3 + CH_3\text{—}HSO_4$$

The first reaction is reversible; thus, in order to drive it to completion, the acid end-product is removed by the addition of silver oxide.

The methyl ethers are the most common derivatives of this type and can be exemplified by fully methylated D-glucose, in which the hydrogens of all five free hydroxyls, including that of the glucosidic hydroxyl, are substituted by CH_3 groups.



Methyltetra-*O*-methyl-α-D-glucose

The glucosidic methoxyl group of the acetal is easily hydrolyzed with acid; however, the other methoxyls, which are true ethers, are resistant to acid as well as to alkali hydrolysis. The stability of the methoxyl group in these reagents makes the methylated sugars extremely useful in structural investigations. It is possible by appropriate methods to methylate selectively the hydroxyl groups in a monosaccharide. These partially methylated derivatives are utilized as reference compounds in the investigation of the constitution of oligosaccharides and polysaccharides

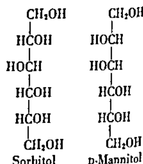
Sugar phosphates. The naturally occurring phosphorylated sugars (D-fructose-1,6-diphosphate, D-fructose-6-phosphate, D-glucose-6-phosphate, α-D-glucose-1-phosphate, D-glyceraldehyde-3-phosphate, and a number of others) are of great metabolic importance. They function as intermediates in the processes of glycolysis, fermentation, photosynthesis, and in most oxidative biological processes. Pentose phosphate esters occur as constituents of nucleic acids and a variety of coenzymes. *See* FERMENTATION; PHOTOSYNTHESIS.

Structurally the phosphorylated sugars are esters. The mono- and diesters of sugars are strongly acidic substances usually isolated as barium, calcium, lead, sodium, cyclohexylammonium, or alkaloid salts. D-Glucose-6-phosphate may be prepared by treating the 1,2,3,4-tetra-*O*-acetyl-β-D-glucose with diphenylchlorophosphonate. The phenyl groups are removed from the resulting tet... *O*-

...rly prepared

...ing the poly-*O*-acetylglycosyl bromide with trisilver phosphate or silver diphenylphosphate. The resulting phosphate triester is simultaneously deacetylated and hydrolyzed under controlled conditions to give the aldose-1-phosphate. By this procedure, α-D-glucose-1-phosphate, α-D-galactose-1-phosphate, α-D-mannose-1-phosphate, and α-D-xylose-1-phosphate have been prepared. The β anomers may be prepared by reacting the poly-*O*-acetylglycosyl bromide with silver dibenzylphosphate, followed by catalytic hydrogenation to remove the benzyl groups and saponification of the acetyl groups. They also can be synthesized by coupling the poly-*O*-acetylglycosyl bromide with monosilver phosphate as the phosphorylating agent.

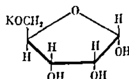
Sugar alcohols. These are acyclic linear polyhydric alcohols. They may be considered sugars in which the aldehydic group of the first carbon atom is reduced to a primary alcohol group. They are classified according to the number of the hydroxyl groups in the molecule. Thus, erythritol with four hydroxyls is considered to be a tetritol; arabitol with five hydroxyls is a pentitol. Best known examples are the two hexitols, sorbitol and mannitol



Sorbitol (D-glucitol, sorbite) is one of the most widespread of all the naturally occurring sugar alcohols. It is found in higher plants, especially in berries and also in algae (seaweeds). Mannitol, like sorbitol, is widespread among plants. However, unlike sorbitol, it is frequently found in exudates of plants.

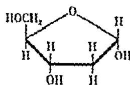
As a group, the sugar alcohols are crystalline substances, having low specific rotations, and ranging in taste from faintly sweet to very sweet. Their distribution in nature is limited to plants.

Ribose; D-ribose. This pentose and 2-deoxy-D-ribose comprise the carbohydrate constituents of nucleic acids which are found in all plant and animal cells. Ribose is soluble in water. It shows complex mutarotation, having a final $[\alpha]_D$ $-25°$ (in water). The best method of preparation is by stepwise hydrolyses of yeast nucleic acid.



α-D-Ribose (furanose configuration)

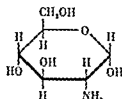
The universal occurrence of D-ribose in all living cells makes this sugar of the greatest interest to biochemists and biologists. Not only is it a constituent of the nucleic acids, but also of several vitamins and coenzymes. The sugar occurs in these natural products in the furanose configuration. *See* NUCLEIC ACID.



2-Deoxy-D-ribose

Deoxyribose (2-deoxy-D-ribose), like D-ribose, is a constituent of nucleic acids in nature. In general, the ribonucleic acids are found in the cytoplasm and the deoxyribonucleic acids in the nuclei of plant and animal cells. Like D-ribose, it occurs in natural products in the furanose configuration.

Amino sugars. These are sugars in which the nonglycosidic hydroxyl groups are replaced by an amino or substituted amino group.



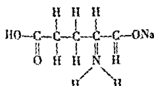
D Glucosamine

The most abundantly occurring amino sugar in nature is D-glucosamine (2-amino-2-deoxy-D-glucose). This amino sugar is a constituent of glycoproteins, which include the mucins from saliva and from egg albumin. It most frequently occurs as chitin, a polymer of *N*-acetyl-D-glucosamine, which is the basis of the hard shells of crustaceans and insects. *N*-Acetyl-D-galactosamine is found in cartilage combined with sulfuric acid and glucuronic acid D-Glucosamine also occurs together with neutral sugars in many bacterial polysaccharides, for example, those of many species of *Pneumococci*. Other types of amino sugars have been recognized in antibiotics. Streptomycin, produced by *Streptomyces griseus*, contains *N*-methyl-L-glucosamine as part of the molecule. A naturally occurring aminopentose, 3-amino-D-ribose, has been found as a component of puromycin (achromycin). Uridine diphosphate *N*-acetyl-D-glucosamine occurs in baker's yeast and in a higher plant (mung bean). *See* PNEUMOCOCCUS; STREPTOMYCIN. [W.Z.H.]

Bibliography: W. W. Pigman, *The Carbohydrates*, rev. ed., 1957.

Monosodium glutamate

The single sodium salt of glutamic acid used in foods to accentuate flavors. It is also known as MSG. Molecular structure is represented as



The crystal form available in commerce is the monohydrate, with structure represented as above plus one molecule of water of hydration.

Glutamic acid is one of the more common of the amino acids. Its structural formula is the same as the monosodium salt excepting that the —CHONa group on the right is replaced by a —COOH group, making the right end similar to the left. *See* GLUTAMIC ACID

The formulas for glutamic acid and its salts show an asymmetric carbon atom. This is the fourth carbon atom from the left It is attached to four entirely different groups. Therefore, the acid itself and each of its salts exist in three forms, the two isomers, L and D, and the racemic or D,L. The L form is the so-called natural or active isomer, and its monosodium salt has the power of bringing out or emphasizing flavors, as distinguished from tastes, of certain foods, notably fish, fowl, meat, and vegetables It is not a flavoring agent but, like salt, aids in developing the savor of foods. As a major constituent of all proteins, glutamic acid participates in many of the metabolic processes.

Source. Originally produced from seaweed in the Orient it is now made principally from cereal glutens, such as those of wheat, corn, and soybeans, and from solutions evolved in the manufacture of beet sugar To be commercially feasible as sources, the proteins from cereals must be concentrated and cheap. The two raw materials used for the greater proportion of commercial production are wheat gluten and desugared beet-sugar molasses. The world's largest single producer is a Japanese firm, principally using wheat gluten. In the United States, five factories exist; two use wheat gluten, one processes corn gluten, and two work solutions from beet-sugar molasses The largest United States manufacturer uses what is commonly termed concentrated Steffen filtrate (CSF). This results from the multiple effect concentration at the sugar factories of the dilute waste liquor produced in recovering sugar from beet molasses by the Steffen process. Another factory uses liquor from both the Steffen and the barium (Deguide) process of sugar recovery.

Glutamic acid appears in the sugar beet as glutamine. During the sugar process, the glutamine changes to the internal anhydride of glutamic acid, pyrrolidone carboxylic acid. The latter is readily hydrolyzed by heating with either acid or alkali, and since it is not a protein, the glutamic acid which results is in the desired L form.

Although basically simple, all of the processes used are somewhat complex because of the many organic substances present in the raw materials. However, commercial production attains a high degree of purity of product, over 99.9% monosodium glutamate

Glutamic acid is also produced by a microbiological method using a carbohydrate as raw material. A United States pharmaceutical manufacturer and a Japanese firm have begun to use this process.

Synthesis. A number of methods for synthesizing glutamic acid from several raw materials

been published in the scientific literature. Synthesis, however, invariably results in the racemized or D,L form. Methods for the resolution of this are known, but the few which have been made public are complex and costly. It has been stated that commercially feasible methods of synthesis and resolution have been developed by two United States chemical manufacturers, one of which is already a large producer of monosodium glutamate. Details have not so far been published.

Uses. United States production includes glutamic acid, its hydrochloride, the mono- salts of sodium, potassium, ammonium, and calcium glutamates, all in the L forms. These find uses in medicine. By far the greatest proportion of glutamic acid production is used as raw material in making monosodium glutamate for the food industry. It is used both in the processing of foods and in the institutional and domestic fields. The Food and Drug Administration has approved it as an additive, and since in itself it is not a flavoring agent, its label designation in use is as a constituent and not as an artificial flavor. Originally, the latter was required because the product made then had a meat-like flavor because of impurities.

Monosodium glutamate is recognized as a standard of identity ingredient in several food preparations now being marketed.

Its principal use is in the preparation of canned and dried soups, but it also enters into the production of some meat, vegetable, fowl, and fish products. It is the "secret" ingredient used by many of the famous restaurant and hotel chefs.

Monosodium glutamate has been and is of great importance in the Oriental diet, and its use as a table condiment in the United States and elsewhere is rapidly growing. In the Orient, it has even been used as a medium of exchange and is often diluted with lactose or salt in order that it may be sold at a price within reach of the poor. See FOOD ENGINEERING; TASTE. [P.D.V.M.]

Bibliography: Armed Forces Quartermaster Food and Container Institute, *Proceedings of Monosodium Glutamate Symposium*, 1948; L. R. Hac, M. L. Long and M. J. Blish, The occurrence of free L-glutamic acid in various foods, *Food Technol.*, 3(10):351–354, 1949; International Minerals and Chemical Corporation, *The Present Nutritional Status of Glutamic Acid*, 1950; P. G. Manning, M.S.G., savoring agent, made in special chemical plant, *Food Inds.*, 20:510–515, 1948.

Monotremata

An order of mammals including the platypus, *Ornithorhynchus*, and the echidnas, *Tachyglossus* of Australia and *Zaglossus* of New Guinea. The monotremes are mammals by definition, because they are covered with hair and nourish their young with milk, but they exhibit many remarkably primitive features otherwise unknown among living mammals. In both the platypus and the echidnas extreme specialization for feeding and locomotion is superimposed on this primitive pattern of organization.

The skeleton exhibits numerous reptilian characters. The skull is extremely modified but retains many reptilian features. There are true ribs on the cervical vertebrae; the shoulder girdle is extremely primitive, with persistent interclavicle, large coracoids and procoracoids, and no true scapular spine. Teeth are transitorily present in *Ornithorhynchus*, these are shed before the animal becomes adult and are replaced by horny pads. The echidnas are entirely toothless. There are no external ears. The urogenital canal is thoroughly reptilian, with a common cloaca into which both the urogenital system and the rectum open. The ureters open directly into the urogenital canal, and oviducts open separately into this canal without forming a true uterus. The eggs are large-yolked, provided with a shell, and are hatched outside the mother's body. The mammary glands are primitive, the individual glands opening independently in a glandular field instead of uniting to form a nipple.

The platypus is semiaquatic, living in burrows along the banks of rivers and feeding on prawns, worms, aquatic insects, and crushable mollusks. The echidnas live in open forest, scrub lands, and rocky areas, and feed on ants and other small insects. See MAMMALIA. [D.D.D.]

Monotremata fossils

The most primitive living mammals are *Ornithorhynchus*, the Australian duckbill, and *Tachyglossus* and *Zaglossus*, the spiny anteaters of Australia and New Guinea, which are known from the Pleistocene through fossils, mostly limb bones. Although monotremes have hair, a squamosal dentary jaw articulation and primitive lacteal glands, they have a reptilelike shoulder girdle and are oviparous. The bill of *Ornithorhynchus* and the mode of modification of the limbs for swimming or digging are unique. The pre-Pleistocene history of the group is unknown, but a recent discovery in England indicates that the monotremes may be related to the Mesozoic docodonts. See DOCODONTA. [W.A.CL.]

Monsoon

Large-scale wind system which predominates or strongly influences the climate of large regions, and in which the direction of the wind flow reverses from winter to summer. The outstanding example of a monsoon circulation is the vast wind system which dominates the weather over most of the Asian continent, especially the southern and eastern parts, the north Indian Ocean, and the extreme western portion of the north Pacific Ocean. The monsoon influence is felt more or less distinctly on all the continental land masses outside equatorial latitudes except Antarctica. The monsoonal circulation in the lower troposphere is cyclonic around the margins of a continental area in summer, and anticyclonic in winter.

General development. These seasonally reversing circulations are attributable to differences in heating or cooling of the atmosphere over continental land masses in contrast to that over the surrounding oceanic areas.

The monsoon influence is absent or feeble in equatorial regions because of the small differences in atmospheric heating from season to season as well as between the cloudy land and oceanic regions. In Antarctica the monsoon influence is not felt because the presence of a permanent ice and snow cover precludes the possibility of relative heating of the atmosphere over the continent even in summer. In many other regions, the monsoon circulation is effectively superposed upon the zonal wind systems and serves mainly to distort them.

Great Asiatic monsoons. This Asiatic system dominates the continent south of approximately latitude 45°. In winter the anticyclonic circulation is centered, on the average, in Siberia near 50°N and 100°E. The mean pressure in this area in the month of January exceeds 30.40 in. (reduced to sea level), the highest such value to be found on the earth. Persistent northerly winds flow from the interior to produce bitterly cold weather far south along the east coast of Asia. When the Asiatic anticyclone is displaced westward from its average position, easterly or southeasterly winds bring outbreaks of cold air from the continental interior into eastern Europe and Scandinavia, and occasionally as far as the British Isles. Prevailing southwesterly flow on the northern and northwestern flanks of the anticyclone produces a relatively mild winter climate in these coastal portions of Asia. Flow of cold air from the north into India and adjacent regions is blocked by the massive mountain barrier along their northern boundaries. At elevations comparable to that of the Tibetan plateau the prevailing wind is part of the circumpolar upper-level westerly circulation. A branch of this circulation descends the mountain slopes and moves across the southern regions in winter as a relatively mild but quite persistent northeasterly wind.

Winter precipitation is sparce in most regions dominated by the Asiatic monsoon circulation because of the small water vapor content of the continental air mass and because of the predominant sinking motion of the atmosphere. Precipitation is plentiful, however, in the East Indies and on the islands of the north Indian Ocean, because the northeasterly wind current experiences a long trajectory over a warm oceanic surface before arriving at these locations with a considerably enhanced water vapor content. This wind current is so firmly established that the intertropical convergence zone is driven well south of the equator. Along this zone rainfall is abundant.

In summer the pattern of the Asiatic monsoon circulation is cyclonic. It is centered about a region of low pressure extending from Arabia to India between latitudes 25 and 30°N. In this region in July the mean pressure reduced to sea level is near 29.50 in. The intertropical convergence zone is displaced far to the north, as a warm and extremely moist southwesterly or southerly wind current prevails over the North Indian Ocean and the extreme western portion of the North Pacific Ocean. Heavy precipitation falls as this current encounters the rugged terrain of southern and eastern Asia. The world's heaviest annual rainfall occurs on the southern slopes of the Himalayas, where, locally, amounts in excess of 400 in., on the average, fall entirely during the summer season. The southwest monsoon begins abruptly on the Indian Peninsula in May or June. In a particular year, the time of onset of the southwest monsoon and the amount of rain that falls may have a crucial effect upon the agriculture of these regions.

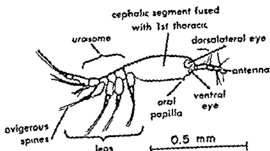
Another portion of the Asiatic summer monsoon system is a persistent northerly wind throughout much of the Near and Middle East, the eastern Mediterranean Sea and extreme southeastern Europe. This current traverses predominantly continental regions and is therefore associated with extremely hot, dry weather.

Effect on North America. On the other continents the monsoon influence is less clear-cut. In North America, for example, monsoon influences include the persistent northwesterly winds in summer along the coast of California; the penetration of warm, moist air from the Gulf of Mexico far to the north in the central and eastern United States during summer with associated copious precipitation amounts; the southward penetration of cold dry air masses in winter in these same regions; and a shift of the prevailing wind from southwest in summer to northwest in winter along much of the Atlantic seaboard. *See* AIR MASS; PRECIPITATION (METEOROLOGY); WIND [F.S.]

Monstrilloida

A suborder of microscopic crustaceans, the least frequently encountered among the subclass Copepoda. They are given the rank of an order by some carcinologists. Many of the 35 species are known only from single specimens, and little is known of their distribution. The life history is unusual. After a free-swimming nauplius larval stage, the immature forms are parasitic upon marine invertebrates. Adults emerge from the hosts, and again become free-swimming in the plankton. Adults lack a second antennae and mouthparts, and the digestive tract is vestigial. Phylogenetic affinities are obscure, but an affinity to the Cyclopoida is suggested. *See* COPEPODA; CRUSTACEA; CYCLOPOIDA. [C.C.D.]

Bibliography: C. C. Davis, A preliminary revision of the Monstrilloida, *Trans. Am. Microscop. Soc.*, 68:245–255, 1949.



Monstrilla reticulata, female. Lateral view. (From Davis, 1949)

Monte Carlo method

A technique for estimating the solution, x, of a numerical mathematical problem by means of an artificial sampling experiment. The estimate is usually given as the average value, in a sample, of some statistic whose mathematical expectation is equal to x. In most of the useful applications, the mathematical problem itself arises in a problem of probability, either in physics or in operational research.

The importance of the method arises primarily from two sources: the practical need to solve equations that are too complicated to solve by analytic methods alone, and the increased importance of all numerical methods because of the advent of the electronic computer.

The method as described above is identical with the earlier method known as artificial or model sampling, or simulation (*see* SAMPLING TECHNIQUES). In fact the name simulation is most appropriate when the mathematical problem arises out of a mathematical model of a real-world situation. The main justification for the name Monte Carlo is that during the 1950s several tricks were introduced for improving the efficiency of the method, so that the subject has assumed a new flavor.

One of the earliest examples of the use of artificial sampling was the experimental estimation of π by the French naturalist, G. L. L. Buffon in 1773. The method is to throw a needle on a striped tablecloth and see how often it falls touching more than one stripe. If the width of each stripe is equal to the length of the needle, then the proportion of "successes" will be close to $2/\pi$ for a long series of trials.

Classification of Monte Carlo methods. The usual method of applied mathematics is to replace a physical problem P by a mathematical problem M, by assuming an adequate mathematical model, to solve the mathematical problem, and thus to solve P (perhaps only approximately). However, sometimes the mathematical problem, when replaced by a numerical problem, is solved with the aid of a calculating machine, so that M is replaced by a new physical problem, P'. One may think of the matter in this way especially if the calculating machine is a so-called analog machine, that is, a machine using nonradix arithmetic (*see* ANALOG COMPUTER). The method then becomes $P \to M \to P'$, where the arrow means "is replaced by." If $P = P'$, it can be said that P is solved by the crude method, or direct experimental method, in which M is inessential.

A special type of experiment with physical apparatus is a statistical experiment S, which makes use of games of chance or their equivalent, such as throwing dice, coins, or even needles. In practice, random sampling numbers are usually used, that is, a sequence of digits generated in such a manner that each selection has an independent chance of 0.1 of giving each of the digits 0, 1, 2, . . . , 9. If a distinction is made between physical and other statistical experiments, there are, among others, the following methods of solving problems: $S \to S$, $S \to S'$, $S \to M$, $M \to S$, $S \to M \to S'$. The method $S \to S$ is the method of estimating the solution of a real-life statistical problem by direct sampling. It may be regarded as a crude form of the Monte Carlo method. In the method $S \to S'$, a real life statistical problem is replaced by a simpler [illegible] which estimation is made by sampling

cation of mathematical statistics. The method $M \to S$ is exemplified by Buffon's needle problem and was suggested for more serious applications by John von Neumann and S. Ulam in 1948. At that time it might have been reasonably called the true Monte Carlo method, but this name is no longer appropriate, because of change of emphasis. It is the method $S \to M \to S'$ which has been of most interest recently. This method can be of varying degrees of sophistication, depending on the amount of mathematical ingenuity required in order to transform S into S' via M.

Advantages and disadvantages. The main advantage of Monte Carlo is that many numerical problems are in practice too complicated to solve in any other method. A familiar example is the estimation of the probability of winning a game of pure chance: sometimes the only reasonably simple method of estimation is to play the game several times. There are also numerical problems that can be solved by deterministic methods but are logically simpler to solve approximately by the Monte Carlo method. The work can then be handed over to unskilled workers with some economy. Sometimes poor approximations are satisfactory because the aim is merely to determine the strategic variables of a problem. This is likely to be a fruitful technique in mathematical economics. To find out how to simplify a complicated economic model without losing touch with reality, one could carry out an approximate Monte Carlo solution for a complicated model, find out which of the variables are important, and then perhaps attempt a mathematical solution using only the important variables.

Another situation where a poor approximation is satisfactory occurs when there is available an iterative method of calculation, that is, a method of successive approximation, which converges to the right answer in a reasonable time provided that the first trial solution is not too far from the truth. The Monte Carlo method may then perhaps be used for obtaining a first trial solution.

The main disadvantage of the Monte Carlo method is that for each extra decimal place required, it is necessary to multiply the sample size by 100. To calculate π to five decimal places by throwing a needle would require about 10^{10} throws, by which time the tablecloth would be worn out.

An advantage claimed for Monte Carlo methods is that they do not become much more complicated when the physical dimensionality of a problem is

increased. They therefore offer some hope for obtaining qualitative information about the solution of partial differential equations in three or more dimensions and about the values of multidimensional definite integrals.

Applications. The method has been applied to the following problems, among others:

1. Size of cosmic ray showers.
2. Critical size of nuclear reactors.
3. Other neutron transport problems, concerning, for example, the shielding properties of water or graphite. The probability that a neutron will cause a branching process that penetrates a shield may be as low as 10^{-10}. In a sample of random walks of reasonable size there will be no "successes" unless some tricks are used.
4. Enumeration of high-polymer molecules or number of self-avoiding walks on a diamond lattice. See LATTICE (MATHEMATICS).
5. Percolation of a liquid through a solid.
6. Brownian motion and diffusion.
7. The birth-and-death branching stochastic process.
8. Autoregressive time series.
9. Theory of queues, and other problems of commercial importance, such as storage, equipment replacement and maintenance, and insurance problems.
10. Laplace's partial differential equation.
11. Schrödinger's partial differential equation
12. Integral equations.
13. Inversion of matrices.
14. Evaluation of definite integrals.
15. Random rounding off.
16. Discovery of the t-distribution by Student

Techniques. When a Monte Carlo method is performed on an electronic computer, the random sampling numbers may be replaced conveniently by deterministic pseudo-random sampling numbers. To do so has the advantage that every job can be checked precisely. Having generated pseudo-random or random sampling numbers, it is trivial to produce random variables that are uniformly distributed in an interval. There are also methods for producing random variables having other probability distributions. Other techniques are concerned with the reduction of the variance of estimates. Example 3 above shows the need for such techniques. If x is a random variable with probability density function $p(x)$, the mathematical expectation of a function $f(x)$, may be estimated by replacing f and p by new functions, f^* and p^*, in such a way that $f^*(x)p^*(x) = f(x)p(x)$, where the integral of $p^*(x)$ is unity (so that p^* can be regarded as a probability point function of a new random variable). It is now adequate to estimate the expectation of f^* for the new random variable. This technique is useful if $f(x)$ is negligible (or zero) except when $p(x)$ is very small. Some other techniques are familiar in sampling survey work, such as the use of statistical regression and stratified sampling. See OPERATIONS RESEARCH; PROBABILITY; PROBABILITY IN PHYSICS; QUEUEING THEORY; STATISTICS; STOCHASTIC PROCESS. [I.J.G.]

Bibliography: E. D. Cashwell and C. J. Everitt, *A Practical Manual on the Monte Carlo Method for Random Walk Problems*, 1959; H. A. Meyer (ed.), *Symposium on Monte Carlo Methods*, 1956; National Bureau of Standards, *Monte Carlo Method*, 1951.

Month

Any of several units of time based on the revolution of the Moon around Earth.

The calendar month is one of the 12 arbitrary periods into which the calendar year is divided (see CALENDAR).

The synodic month is the average period of revolution of the Moon with respect to the Sun, the same as the average interval between successive full moons. Its duration is 29.531 days.

The tropical month is the period required for the mean longitude of the Moon to increase 360°, or 27.322 days.

The sidereal month, 7 sec longer than the tropical month, is the average period of revolution of the Moon with respect to a fixed direction in space.

The anomalistic month, 27.555 days in duration, is the average interval between closest approaches of the Moon to Earth.

The nodical month, 27.212 days in duration, is the average interval between successive northward passages of the Moon across the ecliptic. See TIME. [G.M.C.]

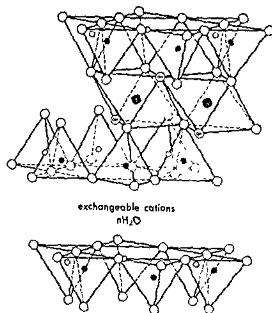
Montmorillonite

A group name for all clay minerals with an expanding structure, except vermiculite, and also a specific mineral name for the high alumina end member of the group. See CLAY MINERALS; VERMICULITE.

Montmorillonite clays have wide commercial use. The high colloidal, plastic, and binding properties make them especially in demand for bonding molding sands and for oil-well drilling muds. They are also widely used to decolorize oils and as a source of petroleum cracking catalysts. See CLAY, COMMERCIAL.

Structure. Because of the extremely small particle size of the montmorillonite minerals, there is still some uncertainty regarding details of their structure. According to the structural concept which is currently accepted, montmorillonite is composed of units made up of two silica tetrahedral sheets with a central alumina octahedral sheet. The atoms in these layers which are common to both sheets become O instead of OH. Montmorillonite is thus referred to as a 3-layer clay mineral with tetrahedral-octahedral-tetrahedral layers comprising the structural unit (see illustration).

These silica-alumina-silica units are continuous in the a and b crystallographic directions and are stacked one above the other in the c direction. In the stacking of these units, the oxygen layers of neighboring units are adjacent. This causes a very



O oxygens ⊕ hydroxyls ● aluminum, iron, magnesium
O and • silicon, occasionally aluminum

Diagrammatic sketch of the structure of montmorillonite. (From R. E. Grim, *Clay Mineralogy*, McGraw-Hill, 1953)

weak bond and an excellent cleavage between the units. Water and other polar molecules can enter between the unit layers and cause an expansion of the structure in the *c* direction. Thus, montmorillonite does *not* have a fixed *c*-axis dimension, but can vary considerably depending on the absence or presence of interlayer molecules. The *c*-axis spacing also varies with the nature of the interlayer cation present between the silicate layers. A montmorillonite in an air-dried condition with sodium as the exchange ion frequently has one molecular water layer and a *c*-axis spacing of about 12.5 angstrom units. Under similar conditions there are two molecular water layers with calcium, giving a *c*-axis spacing of about 15.5 angstrom units. The expansion properties of montmorillonite are reversible; however, reexpansion may be difficult after complete structural collapse by removal of all interlayer polar molecules.

Atomic substitution. The theoretical formula for montmorillonite without structural substitutions is $(OH)_4Si_8Al_4O_{20}\cdot nH_2O$ (interlayer). However, montmorillonite always differs from the above theoretical formula because of structural substitution. In the tetrahedral sheet, aluminum and possibly phosphorus substitute for silicon; [illegible]

[illegible] uioctahedral.

The montmorillonite structure is always unbalanced by the substitutions noted above. The resulting positive net charge deficiency is balanced by exchangeable cations adsorbed between the unit layers and around their edges. The cation-exchange capacity of montmorillonite is normally quite high (100± milliequivalents per 100 grams) and is not appreciably affected by particle size. Substitutions within the structure cause about 80% of the total exchange capacity, and broken bonds are responsible for the remainder.

Other properties. Montmorillonite particles are extremely small and may further disperse in water to units approaching single cell layer dimensions. Most montmorillonite units are equidimensional flakes. However, nontronite tends to occur in elongate lath-shaped units, and hectorite, the flourine-bearing magnesium-rich montmorillonite, is found in thin laths.

There is general agreement that the adsorbed interlayer water between the silicate layers has some sort of definite configuration, but the precise nature of this configuration is not agreed upon. The extent and nature of the orientation of the adsorbed water varies with identity of the adsorbed cations.

When montmorillonite is dehydrated, the interlayer water is lost at a relatively low temperature (100–200°C). The loss of structural (OH) water begins gradually at about 450–500°C, ending at 600–750°C. These temperatures vary with the type and amount of structural substitution. The structure of montmorillonite usually persists to temperatures of the order of 800–900°C. On further heating a variety of phases form, such as mullite, cristobalite, and cordierite, depending on the composition and structure prior to fusion at 1000–1500°C.

Organic ionic compounds enter into cation exchange reactions with montmorillonite. Polar organic compounds, like glycerol, react by replacing the interlayer water causing a shift in the *c*-axis spacing of the montmorillonite units. Thus, the identification of montmorillonite by x-ray diffraction is greatly simplified by preliminary treatment with certain organic reagents. The reaction of montmorillonite and organic material is the base of considerable economic use of montmorillonite clays.

Occurrence. Members of the montmorillonite group of clay minerals vary greatly in their modes of formation. Alkaline conditions and the presence of magnesium particularly favor the formation of these minerals. Montmorillonites are stable over a wide temperature range and have formed by low-temperature hydrothermal processes as well as by weathering processes. Several important modes of occurrence are in soils, in bentonites, in mineral veins, in marine shales, and as alteration products of other minerals. Recent sediments have a fairly high montmorillonite content. *See* BENTONITE; MARINE SEDIMENTS. [F.M.W.; R.F.GR.]

Monzonite

A phaneritic (visibly crystalline) plutonic rock composed chiefly of sodic plagioclase (oligoclase or andesine) and alkali feldspar (microcline, orthoclase, usually perthitic) with subordinate amounts

of dark-colored (mafic) minerals (biotite, amphibole, or pyroxene). Monzonite is more or less intermediate between syenite and diorite. Plagioclase is dominant over alkali feldspar in monzonite but is subordinate to alkali feldspar in syenite. Diorite contains little or no alkali feldspar. See SYENITE.

[C.A.C.A.]

Moon

The natural satellite of the Earth (Fig. 1). The Moon accompanies the Earth in its nearly circular annual sweep around the Sun. It does so at an average distance from the Earth of 384,400 km or 238,860 miles; or, in cosmic terms, at 1/390 of the distance from the Sun, or 60.27 Earth radii. Seen from the center of the planetary system, the Earth would look like a star of −4 magnitude (about as bright as Venus appears from the Earth), of slightly bluish color. The Moon, on the other hand, would look like a star of magnitude +0.5 and be somewhat yellow, like Arcturus. The pair would appear to oscillate in separation like a pendulum with maximum separation of 9 min of arc, two-thirds of the distance between Alcor and Mizar in the Big Dipper, with conjunction occurring every 2 weeks. Seen from the center of the solar system (if one could ignore the glare of the Sun), this binary planet would be the most intriguing and beautiful object in the sky.

Data and dynamical properties. The diameter of the Moon is 3476 km or 2160 miles; its mass 0.012289 Earth masses or 7.343×10^{25} g or 7×10^{19} tons. Its orbit around the Earth is nearly circular, being an ellipse with eccentricity 0.0549. The orbit is described in a sidereal period of $27^d7^h43^m$, measured with respect to the stars or in a nonrotating frame of reference. In terms of the phases of the Moon, which are related to the direc-

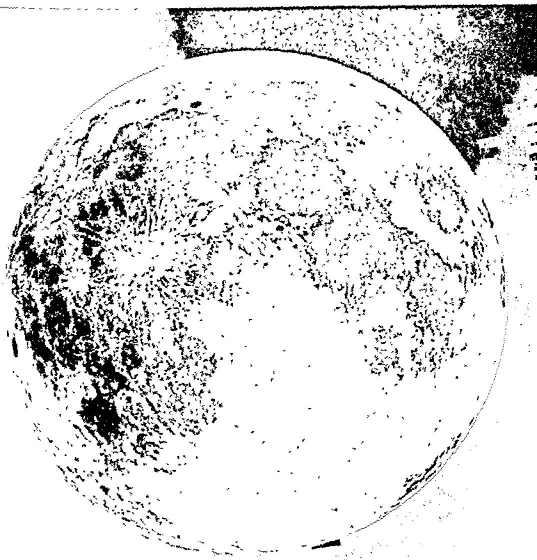


Fig. 1 Full Moon, showing maria (dark), terrae (bright), and ray craters

tion of the Sun and thus change in relation to the stars owing to the motion of the Earth around the Sun, the synodic period, measured with reference to the Sun, is $29^d12^h44^m$ on the average. Because the lunar orbit is eccentric, the length of the synodic month is not constant but varies some 13 hours. The mean motion between the stars is 360°/27.32166 or 13°11′ a day, or about 0.5°/hour. Because the apparent diameter of the Moon is about 31′, the Moon moves approximately its own diameter each hour; its linear speed around the Earth is about 1 km/sec. This is small compared to the common motion around the Sun, which equals 30 km/sec. In fact, if the lunar orbit with respect to the Sun is drawn to scale, the oscillation around the Earth is small enough to leave this larger orbit curved toward the Sun at all times, although the curvature is variable

Lunar motions. The Moon's orbit around the Earth lies in a plane inclined to the plane of the Earth's orbit around the Sun (the ecliptic). The average inclination is 5°8′; its extreme limits are 4°59′ and 5°18′ The lunar orbit intersects the ecliptic in the line of the nodes. This line is important in the occurrence of solar or lunar eclipses. Because the angular diameters from the Earth of the Sun and the Moon are about the same, with both slightly variable owing to the eccentricities in the two orbits, total or near-total (annular) solar eclipses can occur, but only if the Moon is near one of its two nodes and the Sun happens to be there at the same time. Lunar eclipses require Sun and Moon to be simultaneously near opposite nodes. The line of the nodes does not remain constant in space, but regresses in a period of 18.60 years, on the average, owing to differential attraction (tidal force) exerted by the Sun within the Earth-Moon system. This period is reflected in the occurrence of eclipses. Also under the influence of this tidal force, the major axis of the lunar orbit moves, in its own plane, with a period of 8 850 years, in a forward direction; that is, the line of apsides advances with this period.

The rotation of the Moon is synchronized with the Moon's motion in its orbit so that approximately the same face is always presented to the Earth. However, because the rotation is nearly uniform but the motion in the orbit is not, owing to the finite eccentricity and Kepler's law of areas, the face of the Moon turns some 8° right and left from the central position during each monthly period. This apparent wobble is the libration in longitude. There is also a libration in latitude because the plane of the lunar equator makes an angle with the plane of the orbit of about 6°40′. These librations make it possible, by stereoscopic methods, to derive elevation differences on the Moon, although the precision is limited to 1000–2000 ft, because of the great distance of the Moon from the Earth. More accurate elevation differences can be found from shadow measurements and a knowledge of the position of the Sun and the lunar features. Both methods must be used when the entire lunar figure is to be measured.

The fact that the Moon always turns the same face toward the Earth, apart from the small librations, is the first of three laws formulated by J. D. Cassini of Paris in 1692 describing the observed rotational properties. These laws are (1) The Moon rotates uniformly about an axis which is fixed with respect to the Moon itself; the period of this rotation is identical with the sidereal period of the Moon in its orbit, namely 27.321661 days. (2) The pole of the lunar rotation z makes a constant angle (1°35′) with the pole of the ecliptic Z, which may here be regarded as a fixed point on the celestial sphere. (3) In consequence of the nearly uniform regression of the lunar node on the plane of the ecliptic and the nearly constant inclination of the lunar orbit (5°9′), the pole of the Moon's orbit P is known to describe a small circle about Z in a period of 18⅔ years. The arc of a great circle zP contains also the pole Z. In other words, the planes of the lunar orbit and the lunar equator intersect on the ecliptic, the latter plane being intermediate between the two former.

Lunar shape. The dynamical interpretation of these laws, by C. J. L. Lagrange, P. S. Laplace, and S. D. Poisson, has given important information on the differences between the three principal moments of inertia, and by inference, on the shape of the Moon. The differences between the three axes of the Moon regarded as a triaxial ellipsoid are derived in this way with much more accuracy than by direct measurement, but only if the mass distribution along the three axes is basically the same. On this plausible assumption it is found that the radius pointing toward the Earth is 1.1 km longer than the polar radius. It follows from theory that the percentual precision of this quantity is the same as that of the tilt of the lunar equator on the ecliptic, or about 1%. Direct measures of the difference based on plates taken at different librations have given larger values, up to 5.7 km, but these are suspected to contain systematic errors of measurement. The third lunar axis, along the direction of motion around the Earth, is intermediate in size.

One important question is concerned with whether the length of the axis pointing toward the Earth is the result of a fossil (or frozen) tidal bulge. If this were so, the distance from the Earth at which the bulge was frozen can be computed dynamically from the difference between the polar and radial axes; 0.39 of the present lunar distance is found. This value is not implausible on other grounds, but the argument itself is weakened by the fact that the third, intermediate axis does not have quite the required length for the Moon to have been an equilibrium figure at any distance from the Earth. Apparently the tectonic irregularities on the Moon's surface are sufficiently large to affect the moments of inertia measurably. This, however, leads to another conclusion of vital importance, namely that the major tectonic features on the Moon, described below, are not of volcanic origin. If they were, the symmetry properties of the Moon would not have been appreciably disturbed. Instead, these features must be the result of major

impacts that added mass locally or disturbed the symmetry by large displacements. The Moon has been a nonequilibrium figure for an estimated 4,000,000,000 years. If the Moon's *body has yielded* partially in the course of time, its plastic condition will have occurred even closer to the Earth than is estimated from the present bulge.

Earth-Moon interaction. The Moon causes tides in the oceans which follow the Moon around the Earth with a phase lag caused by frictional forces acting largely in shallow waters. Because of this phase lag, the Moon exerts *a couple on the Earth*, slowing down its rotation, whereas conversely, the tidal bulges exert a couple *on the Moon, tending* to accelerate it in its orbit. The dual result is that the Earth day *increases in length and that the* distance from Moon to Earth *increases, thereby* increasing the length of the lunar month. In past ages, the Moon must *have been closer and the day shorter*. The present rates are such that the Moon may have been *quite close to the Earth* about 4,000,000,000 years ago, the approximate age of the Earth-Moon system.

Dynamically *there is a limit beyond which this* process cannot have started, when the day was about 5 hours and the separation was about 3 Earth radii (compared to the *present 60*). *Had the Moon been* closer, it would have broken into fragments by tidal action *of the Earth. There is indirect evidence*, from the abundances of the atmospheric noble gases of *high atomic* weights (krypton and xenon) that the Earth had a rapid rotation initially; otherwise *a larger* fraction of the original supply of these gases would have been retained. (They are too heavy to be lost by evaporation from a slowly *rotating exosphere* near the Earth.) It may be tentatively assumed that the Moon did not form much *beyond 3* Earth *radii* away, and its distance from the Earth [illegible]

Ob [illegible] the Moon [illegible] than 10-13 of Earth's), its surface is visible with complete *crispness* whenever Earth's atmosphere permits observation. This is in accordance with theoretical expectation for this comparatively small body from which any gases released by the crust would rapidly escape into the interplanetary vacuum. The Moon is thus quite unlike Venus, where a dense veil hides the surface, or Mars, where atmospheric haze and dust often impede observation. The absence of a lunar atmosphere prevents erosion of the types most effective on the Earth, by water, ice, and wind. Except for features destroyed by melting or by lava flows and destroyed or damaged by impacts, large or small, the Moon has preserved the surface features acquired through its entire history.

Most of the mass falling on the Moon from interplanetary space is composed of particles from the zodiacal cloud that envelops the Earth, which are a few microns in size and smaller. Most impacts are consequently in the nature of pinpricks. (The zodiacal particles are of cometary origin and are called micrometeorites when they enter the Earth's atmosphere.) In addition, the lunar surface is exposed *to solar* ultraviolet and x-radiations, as well as solar particles, mostly protons and electrons, and to cosmic rays. These various radiations will not *affect the geometry of* the surface but may change some of its optical properties. The Moon, then, presents a fossil record of events in the immediate vicinity *of the Earth*, going back to the earliest history of the Earth-Moon system; it is a precious source of information.

In the modern age of large telescopes, the scale of discovery of surface markings is set essentially by the disturbing effects of the atmosphere above the observatory. Photographically, resolutions down to just under ½ sec of arc (0″.5) have been attained, while visually, under the best conditions, it is possible to reach 0″.1. Because the mean angular diameter of the Moon is 31′5″ or 1865″, and because the linear diameter is 2160 miles, 1″ equals 1.16 miles. The photographic *resolution is* thus about 0.5 mile, and the best visual resolution 0.1 mile. The vertical [illegible]

high may be seen because when the Sun *is only 1°* above the horizon, the shadow will be 0.1 mile wide.

The varying phases of the Moon allow observation of the surface under very different illuminations. At full Moon no shadows are seen and many familar features disappear from view entirely whereas others, such as the ray systems, attain maximum visibility. At the first and last *quarters* the terminator crosses the center of the disk, and nearby regions are seen under oblique illumination, greatly enhancing the appearance of relief. However, the long shadows hide much detail and observations at less oblique illuminations must be added to fill the gaps. Thus the complete picture of the lunar surface requires a combination of data from many different phases. Some hasty announcements, of lunar bridges, for example, have resulted from the failure to appreciate this requirement.

Several lunar maps and atlases have been made. The earlier maps were necessarily based on visual observations. The best map produced during the nineteenth century was that by J. Schmidt of Athens, published in 1878 at Berlin. On it the Moon has a diameter of 6 ft; the map is divided into 25 sections, and shows 32,856 lunar craters. It was based on 30 years of visual studies and thousands of sketches, mostly made with the 6-in. Athens telescope. The resolution of a 6-in. telescope is about 0″.75, so that the best of the modern photographs have overtaken the Schmidt map. Furthermore, photographs record all that is on the lunar surface within the resolution attained; completeness can never be obtained on a visually constructed map. Finally, the photograph records with precision. One good photograph shows more than could possibly be drawn with precision in 1,000 hours. Nevertheless, the value of visual observation has not diminished, because under the best conditions it greatly excels photography in resolution. The modern procedure is therefore to obtain all possible detail

photographically and to make all measurements from photographs; visual work supplements and interprets the features shown on the photographs.

An excellent photographic lunar atlas, consisting of 71 large reproductions, was issued by the Paris Observatory about 1900; the scale is not the same for all charts, being 5–10 ft to the lunar diameter. A new photographic lunar atlas is based on the best photographs now in existence, obtained at the Mt. Wilson, Lick, Yerkes, McDonald, and Pic du Midi observatories. It consists of 230 sheets, each 16 by 20 in., with a scale of 100 in. to the lunar diameter. Each lunar area is shown under at least four different illuminations. A supplement will show the lunar limb areas, highly foreshortened on normal photographs, reprojected, with the foreshortening removed.

Surface structures. The surface of the moon shows three types of structure: (1) maria, which are dark areas, reflecting 3–5% of the sunlight received, roughly level, although by close inspection nowhere precisely level or smooth, almost certainly composed of lava basins; (2) highlands (sometimes called terrae, the seventeenth-century designation), used here to refer to the apparently original lunar crust, modified only by impact craters, but not by large-scale melting or a heavy overburden of ejectamenta; and (3) areas that are clearly covered with ejectamenta, to the point where the original surface is almost or completely unrecognizable (Figs. 1 and 2). These loaded areas are found surrounding the impact maria (see below) and some of the larger impact craters (Fig. 3).

All three types of surface show impact craters, but the highlands have a much larger surface frequency than the other regions. This can mean only that the highlands are older, and that the maria



Fig. 2. Names of principal features of the Moon.

and loaded regions show only the postmare impacts. This is a slight over-simplication; some of the maria show ghost craters that appear to be the remnants of walls of premare impact craters; other maria have no ghost craters. This provides one of the criteria used to subdivide the maria into two generic classes: impact maria, which are roughly circular in outline, are surrounded by mountain walls, and have no ghost craters; and those maria presumably caused by lava flooding.



Fig 3. Crater Plato in Lunar Alps, with Alpine Valley (left), Mare Frigoris (center), W. C. Bond and other straight-wall formations (bottom). The straight lines are oriented either to the Imbrium Impact Center, or are at right angles to it. A large rock mass has slumped on inner slope of Plato, right side.

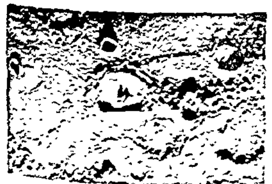


Fig 4. Theophilus and surroundings. Shadow shows profile of central mountains. Rubble tossed out of crater has covered central mountains of Cyrillus (right above), making them rounded and streamlined. Numerous scars from main crater show on lower and left sections.

Tentatively, Maria Crisium, Nectaris, Serenitatis, Imbrium, and Humorum are classified as impact maria (Figs. 1 and 2). Those caused by flooding appear to be *Maria Tranquillitatis, Foecunditatis*, Nubium, Frigoris, and Oceanus Procellarum, as well as several smaller basins, Mare Vaporum, Sinus Medii, Sinus Aestuum, Lacus Somniorum, and Sinus Roris. It is assumed that the impact maria *utterly destroyed the previous crust so that no ghost* craters can be present; flooding would not do this. Mare Imbrium, the largest of the impact maria, shows one property not shared by the others. The impact appears to have caused an enormous forward lava splash that *covered the Apennines and* surrounding areas, whereas the impacts of Maria Crisium, Nectaris, and Humorum appear to have been dry, with no evidence of splashing, and with the lava flooding of the basin occurring later. The formation of Mare Serenitatis appears to have been followed by much melting and subsidence, partly connected with the Imbrium impact, and its own development is therefore partly hidden.

The craters belong to several subclasses which appear to arrange themselves according to age. No absolute ages can be found, but criteria for relative ages exist. First of all, there is good reason to believe that all the maria are approximately of the same age. Craters inside maria are therefore postmare, and craters on the terrae are either premare or postmare Because the terrae have such a dense coverage of large craters, the latter are mostly premare in age

The postmare craters belong to four categories: (1) ray craters, (2) early craters, (3) intermediate-aged craters, and (4) craters with central vents.

Of the dozen or so very conspicuous ray craters (Fig. 1), Tycho, Copernicus, Kepler, Aristarchus, Aristillus, Langrenus, and Proclus are the best known (Fig. 2) Two of these occur on terrae but their rays extend over maria as well. No crater rays from terrae are observed to stop at a mare shore line; thus all ray craters are postmare. The rays of Copernicus particularly contain numerous scars on the lunar surface, many about 1-2 miles long and a few tenths of a mile wide, clearly caused by flying fragments ejected from the crater. Apparently the rays consist of spurts of ejectamenta, containing both finely divided white material (possibly pulverized rock or glass beads) and coarser materials. *Scars are associated only with postmare* craters; among the best examples are Copernicus, Theophilus (Fig. 4), Langrenus, Aristoteles, Aristillus, Eratosthenes, and Tycho. The occurrence of large blocks tossed out by the ray craters shows that at the time of their formation the lunar surface was coherent, composed of igneous or metamorphic rock, not accreted rubble. This is consistent with the assumption that the maria are lava basins and the terrae had consolidated into metamorphic rock [illegible] postmare period The white ma- [illegible] al (rock [illegible] ter rims and the small irregular [illegible] ns char-

acteristic of these craters also seem to indicate a brittle surface.

Some large early postmare craters apparently formed when the lunar surface was still hot. The crater floor melted from the action; examples are Plato (Fig. 3) and Archimedes.

Other craters are apparently intermediate in age between ray craters and early postmare craters, such as Bullialdus, Aristoteles, and Theophilus (Fig. 4), which are surrounded by prominent scars. Their crater rays are weak.

Some fairly large craters appear to have, instead of central peaks, central volcanic vents, which are lined with bright elevated rims showing flow patterns. The principal objects are Timocharis, Lambert, Euler, and Pytheas in Mare Imbrium; and Plinius and Maskelyne in Mare Tranquillitatis. The vent in Timocharis is wide open, nearly square, and about 1 mile in size. In Plinius it is nearly filled in with light-colored flows. The absence of lava bottoms and the near-absence of crater rays seems to place the age of these six craters between the early postmare craters and those of intermediate age.

The postmare time sequence is therefore (1) lava floors, no rays (Fig. 3); (2) volcanic vents, but no general lava floor, and absent or very minor rays; (3) prominent and well-formed central mountains and prominent systems of scars, but only weak rays (Fig 4); (4) minor, irregular central mountains, very bright at full Moon, with a rubble-covered crater floor, bright crater rims at full Moon, prominent scar systems, and bright rays (Fig. 1). Apparently the temperature of the lunar surface at the time of impact controlled the resulting type of crater. The occurrence of prominent central mountains which do not appear to be piles of debris but orderly igneous extrusions seems to require an intermediate temperature, low enough for the displaced crust to have remained solid but high enough for the presence of lavas at or just below the impact center.

In addition to these four groups of large postmare craters there is a fifth group comprising thousands of small craters, usually less than 1 mile across, surrounded by white halos three or four times the diameter of the crater; it is most likely that these are meteorite craters, similar to the one in Arizona, which they approach in size. Mare Serenitatis (Fig. 5) and Mare Humorum are particularly favorable locations for their study; their numbers on the Moon are roughly compatible with those on Earth (where only the most recent members have survived). A sixth group consists of many dozens of craters 1–10 miles in diameter, not accompanied by prominent halos (Fig. 5). These craters are usually extremely regular, bowl-shaped, with the steepest part just inside the rim, having slopes of about 38°, apparently close to the angle of repose. Often their bottoms are not parabolic in shape but somewhat conical, with the bottom part sloping about 15°. The smaller members of this class have a nearly conical appearance (Fig. 6, upper half); the larger ones have a flat dark floor usually at about a constant depth below the surrounding terrain. The rims of these craters are usually low. Although typical bowl-shaped craters on the Moon normally have rims whose volume will approximately fill the depression (Schroeter's rule), this is far from true for the cone craters.



Fig 5. West half of Mare Serenitatis, showing Serpentine Ridge, many small bright spots, each of which contains a small crater, and some larger craters without halos. Ridge shows en échelon structure; that is, it is composed of parallel but displaced segments, a common feature in terrestrial island arcs.

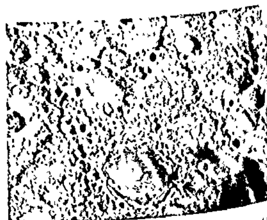


Fig. 6. Shore region south of Mare Nubium Very old craters are in upper half. Pitatus has lava bottom and peripheral tension rill. To the right is a circular squeeze-up on a small crater floor; below it, a graben Cone craters of different dimensions in upper half

The mass may have been blown away with such force that little of it fell near the crater. The number of cone craters per unit area is similar for the maria and the terrae; apparently the age of the cone craters is postmare. Some small cone craters occur in crater chains (Fig. 7, right center), which cannot have been caused by impacts, by the laws of chance, but may be a series of blow holes above a lunar dike.

The terrae contain two types of craters not present in the maria, and therefore are presumably premare. Those of the first type are entirely or nearly invisible at full Moon. Their walls are usually heavily damaged by subsequent impacts (Fig 6, top) and as a class they must be old. They have no or minor central mountains; in many cases the floors are flat and smooth as if they had been raised and leveled, presumably by melting at subsurface levels; as a result, the larger craters in this class, such as Clavius, may have convex floors. Craters of the second type show bright rims at full Moon. These usually have sustained only moderate damage; they have either prominent central mountains on a light-colored floor, or no mountains on a darker floor (melting). Examples are Arzachel and Ptolemaeus (Fig. 7). The latter has a floor containing some 20 shallow round basins of a type that can have been produced only in a viscous material.

It is believed that the first type of crater was formed by impacts in unconsolidated, accreted rubble. Like an impact in sand, the material did not melt, being merely displaced, and hence the crater is invisible at full Moon. The second type apparently formed after the crust had been metamorphosed into a white rock, which according to H. H. Hess, may be one of the magnesium silicates that are all white: enstatite, formed at high temperature, amphibole, at medium temperature; and talc, at low temperature.

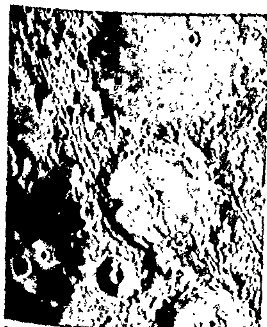


Fig 7. Ptolemaeus (bottom left), Alphonsus (center left), Arzachel (top left, half shown), Alpetragius (with central mountain soft in outline, apparently the result of shower from Alphonsus) Crater chain (right center); grooves and graben of Imbrium system traverse lower right to upper left



Fig 8 Copernicus ejectamenta (left) from impact crater, volcanoes near Hortensius (just above center) with one at extreme right, large, low domes, some with calderas (lower right)

In addition to the maria and the craters, with accompanying flooding and ejectamenta, there are many subsidiary structures on the Moon including tension rills (Fig 6, bottom), often located near the outer borders of a mare; pressure ridges (Fig. 5, left), present on the maria and clearly related structurally to the geometry of the mare basins; faults, such as the Straight Wall, many of which are related to the Imbrium Impact, or the enormous fault scarps along the Apennines and the Altai Mountains; graben, such as the Rheita Valley, many of which are related to the Imbrium Impact, and others, such as the straight graben along the south shore of Mare Nubium (Fig. 6, right bottom), which are related to mare basins; lava domes, a large group of which occurs near Marius; extinct volcanoes, such as the group of six near Hortensius (Fig. 8), objects which have calderas or craters in their tops of about 0.2 of the crater diameter, as well as much larger circular, dome-shaped lava blisters which usually also have a central caldera (Fig. 8); volcanic sinks, such as those between Hortensius and Reinhold (Fig. 8); and squeeze-ups or extrusion walls, such as the several parallel walls near W. C. Bond (Fig. 3), or near



Fig. 9. Far side of Moon photographed by automatic space station. To left of dashed line is a portion of the side of the Moon that faces Earth, the photograph shows such familiar features as I, Humboldt's Sea; II, Sea of Crises; III, Marginal Sea, IV, Sea of Waves; V, Smyth's Sea; VI, Sea of Fertility, and VII, Southern Sea. To right of dashed line are newly discovered features: 1, Moscow Sea; 2, Astronauts' Bay; 3, continuation of South Sea; 4, Tsiolkovsky Crater, 5, Lomonosov Hill; 6, Joliot-Curie Crater; 7, Soviet Mountains and Dream Sea. (Sovfoto)

the northeastern border of the Aristarchus Uplift, or, on a smaller scale, on many pressure ridges (Fig. 5). All in all, the Moon is an extraordinarily interesting object, in which the elemental forces of nature have had free play without the subsequent obliteration that has occurred on the Earth.

Thermal history. Studies of meteorites, which by all indications are fragments of colliding asteroids, show that the asteroids went through a process of melting and subsequent solidification about 4,600,000,000 years ago. There is some evidence that at about the same epoch the Earth also obtained its present main arrangement as a result of melting and subsequent freezing. The Moon, which is intermediate in size, may therefore be expected to have done the same. The maria are therefore about 4,500,000,000 years old and apart from some volcanic activity that seems to have occurred on the Moon soon after the maria froze over, there have been only impacts to change the appearance of the lunar surface.

The maki...

re

so

M... Early in 1960 it was discovered that ... asteroids or the is present in meteorites. This proves that the asteroids were formed not over 300,000,000 years after nucleogenesis and that short-lived radioactive isotopes will have contributed to the radioactive heating process. The same is presumably true for the Moon. Its age is, therefore, assumed to be slightly under 5,000,000,000 years. The numerous and large impacts of the premare period cannot have been asteroids, which initially must have moved in nearly circular orbits between Mars and Jupiter, but rather small satellites encircling Earth. The Moon, in its outward journey resulting from tidal friction, appears to have swept up most of these small bodies, causing craters and impact maria. See ASTEROID; METEOR.

Recent studies. A major advance in lunar studies was made in 1959 by the scientists of the U.S.S.R. with the launching of three successful Moon rockets, Lunik I, II, and III. Two main results have been announced: the measurement of the lunar magnetic field, and the photography of the back of the Moon.

The strength of the magnetic field on the lunar surface was determined by Lunik II, which struck the Moon somewhere between the craters Archimedes and Aristillus, according to the orbit computations. The field was found to be less than 0.0003 gauss, or less than 1/2000 of that on the surface of the Earth. The limit was set not by the sensitivity of the recording device but by fluctuations in the field caused by clouds of ionized solar particles passing through interplanetary space past the Earth and Moon. The absence of a measurable field is interpreted to mean that a liquid iron core (in convective motion) is absent from the Moon. This had been surmised from the low mean density of the Moon, 3.33 g/cm^3, which leaves no room for a sizable iron core; and from the comparatively small size of the Moon, which suggests that its interior has cooled appreciably below the temperatures found in the interior of the Earth. The direct measurement of the magnetic field is a valuable result.

Even more significant is the photography of the back of the Moon. It was carried out by Lunik III on October 7, 1959 during an interval of about 40 minutes after the rocket had passed the Moon and saw about ¾ of the unknown back and ¼ of the known front. Because the Moon was a crescent as seen from the Earth, the rocket had essentially a full-Moon view during the photography and could record only maria but no craters, shadows being absent.

Two cameras were used, of 8- and 20-in. focal length; the first recorded the entire Moon, the second only a segment. Numerous frames were taken and automatically developed in a chamber whose temperature was controlled. Several dozen pictures were successfully transmitted to Earth by a photoelectric scanner; the transmissions were repeated several times, first when the rocket was still beyond the Moon and later when it had returned closer to the Earth. The transmissions were made slowly to reduce the effects of superposed radio noise. Figure 9 shows the features recorded. Those encircled by a full line were present on all of the best 10 frames, and the Soviet Academy has assigned names to them. Features present on 2-5 frames are encircled by dotted lines; they are re-

garded as less certain and no names have been given to them.

The Russian photographs suggest that the back of the Moon has fewer maria than the front; nothing can be said about craters for the reasons stated. In explanation, it has been suggested that the asymmetry may be due to Earth itself. Some small effect could indeed have been produced by impacts hitting the front side preferentially, as appears to be expected dynamically, while another effect may be caused by the fact that one major impact on the front (Mare Imbrium) caused many subsidiary maria to form (Mare Frigons, Mare Nubium, and possibly Oceanus Procellarum). [G.P.K.]

Moose

A large deer, *Alces alces*, found in northern Eurasia and the northern part of North America. The American moose is frequently considered a distinct species, *A. americana*. The moose is the largest of all deer, weighing as much as 1300 lb. It is an animal of the deep woodlands and shallow lakes of the north, and spends much time in shallow water feeding on submerged vegetation. It also eats willows and other water-border plants. The moose can run with surprising speed. The male has large, palmate antlers bordered by numerous small



The moose, *Alces americana*; length to 9 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

points. Its flesh is a staple winter food for many of the peoples inhabiting the northland, but the moose has been exterminated or seriously depleted over much of its original range. *See* ARTIODACTYLA; DEER [J.D.B.]

Moraine

Glacially derived soil materials or drift deposited at end, side, or under a glacier. Moraines may be classified according to topography and position as terminal or end, ground (formed when the ice melted), lateral (on the side of a valley glacier), and medial (where two valley glaciers joined). The material of moraines is chiefly till with some strati-





Rolling, hummocky terrane of terminal moraines. (a) Cut-over second-growth forest area of Wisconsin (*Wisconsin Geological Survey*). (b) Cleared farm land of Pennsylvania (*W. C. Alden, USGS*).

fied deposits. End moraines require moving ice the margin of which was held nearly stationary by a balance between motion and wastage. They are commonly complex ridges parallel to the ice border. Ground moraine owes much of its form to preexisting rock topography. *See* GLACIATED TERRANE. [F.T.T.]

Moraxella lacunata

The cause of a type of chronic conjunctivitis in [illegible] organism has been called the Morax- [illegible] than media containing blood. In diagnostic stained smears from infected eyes the organisms frequently appear in pairs and are somewhat larger (0.5 by 2.0 microns) than the other hemophilic bacteria. Neomycin ointment is often used to treat this infection. *See* HEMOPHILIC BACTERIA; NEOMYCIN. [W.F.V.]

Mordant

A substance or combination of substances that facilitates the fixing of a dye to a fiber. A mordant enables the production of a more permanent and often deeper color. Metallic salts or hydroxides are most *frequently* used as mordants.

Certain mordants act directly on the fiber, making it more susceptible to the dye. Fabrics are then

The details of the life histories of the various species differ greatly. Larvae feed upon plant and animal plankton, adult males upon exposed plant juices and nectar, and most adult females upon vertebrate blood, which they must have before they can mature eggs. A few mosquitoes eat only plant juices and never bite animals. *See* DIPTERA; PARASITOLOGY. [J.D.B.]

Moth

Any of a large number of species of insects belonging to the order Lepidoptera, including both the butterflies and moths. The latter are distinguished by their large bodies and nocturnal habits, in contrast to the day-flying and relatively slender-bodied butterflies. The antennae of moths are varied but usually not knobbed, whereas all butterflies have knobbed antennae. Moths hold their wings in the horizontal position when at rest; the wings of butterflies are folded together over the back.

Characteristics. All of the order have sucking mouthparts (in a few species they are vestigial) as adults, and chewing mouthparts as larvae. The larvae are called caterpillars and almost all are plant eaters. All Lepidoptera have complete metamorphosis, and the caterpillar passes through several molts before entering the resting, or pupal, stage. In contrast to butterflies, most moths pupate in a cocoon, and their pupae are frequently called chrysalids.

Although some of the largest and showiest of our insects are moths, many moths are well-camouflaged, drab insects that frequently escape notice. Moths usually have but one generation a year, and overwinter in the pupal or larva stage. However, some spend the winter as adults, and a few others as eggs.

Best known of the showy moths are the giant [illegible]

[illegible] hawk moths, include a large number of narrow-winged moths which are frequently gray above but show brilliant red, black, yellow, and orange markings in flight. This family also includes the clear-winged sphinx, often mistaken for a hummingbird and commonly called the hummingbird moth. The tobacco or tomato worm is the larva of a large sphinx moth, *Protoparce quinquemaculata*. The larvae of this family are often called hornworms because of the prominent spines borne on the eighth abdominal segment of most species.

Pests. The large family Noctuidae (Phalaenidae) includes some of the most destructive agricultural pests, notably the army worm, cutworm, corn earworm, and cabbage looper. The family Liparidae includes several serious pests of forest and shade trees, such as the white-marked tussock moth and two introduced species, the gypsy moth and brown-tailed moth, both of which have caused extensive damage to trees in the northeastern United States. Control measures undertaken against the last two have been extremely expensive and quite destructive to wildlife.

Other well-known moth pests include the clothes moth, tent caterpillar, cankerworm, and the European corn borer. *See* CANKERWORM; CATERPILLAR; CORN.

Perhaps the only species helpful to man in the entire group is the silkworm moth, *Bombyx mori*, the only species of the family Bombycidae. *See* ARMY WORM; BUTTERFLY; CUTWORM; INSECTA; LEPIDOPTERA; SILK. [J.D.B.]

Motion

If the position of a material system as measured by a particular observer changes with respect to time, that system is said to be in motion with respect to the observer. Absolute motion, then, has no significance, and only relative motion may be defined, what one observer measures to be at rest, another observer in a different frame of reference may regard as being in motion. *See* FRAME OF REFERENCE; RELATIVE MOTION.

The time derivatives of the various coordinates used to specify the system may be used to prescribe the motion at any instant of time. How the motion develops in subsequent instants is then determined by the laws of motion. In classical dynamics it is supposed that in principle the motion and configuration of the system may be specified to an arbitrary precision, although in quantum mechanics it is recognized that the measurement of the one disturbs the other.

For a system of f degrees of freedom, the motion may be represented by a point in an f-dimensional velocity space, the coordinates of which are the time rates of change of the coordinates that describe the configuration of the system. For a system under no forces that is described by rectangular cartesian coordinates, these time derivatives are constants. For a single particle, this result is the first of Newton's laws of motion, namely that a particle remains at rest or in a state of uniform motion unless acted upon by an external force.

The most general theory of motion that has yet been developed is quantum field theory, which combines both quantum mechanics and relativity theory, as well as the experimentally observed fact that elementary particles can be created and annihilated. *See* QUANTUM FIELD THEORY; *see also* DEGREE OF FREEDOM (MECHANICS); DYNAMICS; EULER'S EQUATIONS OF MOTION; HAMILTON'S EQUATIONS OF MOTION; HARMONIC MOTION; KINEMATICS; KINETICS (CLASSICAL MECHANICS); LAGRANGE'S EQUATIONS; MOTION, RECTILINEAR; NEWTON'S LAWS OF MOTION; OSCILLATION; PERIODIC MOTION; QUANTUM MECHANICS; RELATIVITY; ROTATIONAL MOTION. [H.C.C.]

Motion, rectilinear

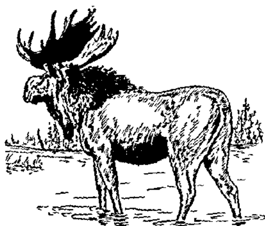
Motion is defined as continuous change of position of a body (*see* MOTION). If the body so moves that every particle of the body follows a straight line

garded as less certain and no names have been given to them.

The Russian photographs suggest that the back of the Moon has fewer maria than the front; nothing can be said about craters for the reasons stated. In explanation, it has been suggested that the asymmetry may be due to Earth itself. Some small effect could indeed have been produced by impacts hitting the front side preferentially, as appears to be expected dynamically, while another effect may be caused by the fact that one major impact on the front (Mare Imbrium) caused many subsidiary maria to form (Mare Frigons, Mare Nubium, and possibly Oceanus Procellarum). [C.P.K.]

Moose

A large deer, *Alces alces*, found in northern Eurasia and the northern part of North America. The American moose is frequently considered a distinct species, *A. americana*. The moose is the largest of all deer, weighing as much as 1300 lb. It is an animal of the deep woodlands and shallow lakes of the north, and spends much time in shallow water feeding on submerged vegetation. It also eats willows and other water-border plants. The moose can run with surprising speed. The male has large, palmate antlers bordered by numerous small points. Its flesh is a staple winter food for many of the peoples inhabiting the northland, but the moose has been exterminated or seriously depleted over much of its original range. *See* ARTIODACTYLA; DEER. [J.D.B.]



The moose, *Alces americana*; length to 9 ft. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Moraine

Glacially derived soil materials or drift deposited at end, side, or under a glacier. Moraines may be classified according to topography and position as terminal or end, ground (formed when the ice melted), lateral (on the side of a valley glacier), and medial (where two valley glaciers joined). The material of moraines is chiefly till with some stratified deposits. End moraines require moving ice the margin of which was held nearly stationary by a balance between motion and wastage. They are commonly complex ridges parallel to the ice border. Ground moraine owes much of its form to preexisting rock topography. *See* GLACIATED TERRANE. [F.T.T.]





Rolling, hummocky terrane of terminal moraines. (a) Cut-over second-growth forest area of Wisconsin (Wisconsin Geological Survey). (b) Cleared farm land of Pennsylvania (W. C. Alden, USGS).

Moraxella lacunata

The cause of a type of chronic conjunctivitis in humans. This organism has been called the Morax-Axenfeld bacillus and *Haemophilus duplex*. Loeffler's serum medium grows this organism better than media containing blood. In diagnostic stained smears from infected eyes the organisms frequently appear in pairs and are somewhat larger (0.5 by 2.0 microns) than the other hemophilic bacteria. Neomycin ointment is often used to treat this infection. *See* HEMOPHILIC BACTERIA; NEOMYCIN. [W.F.V.]

Mordant

A substance or combination of substances that facilitates the fixing of a dye to a fiber. A mordant enables the production of a more permanent and often deeper color. Metallic salts or hydroxides are most frequently used as mordants.

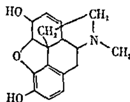
Certain mordants act directly on the fi
ing it more susceptible to the dye. Fabri

pretreated with the mordant before exposure to the dye. For example, cotton is soaked in a mixture of aluminum sulfate, $Al_2(SO_4)_3$; sodium carbonate, Na_2CO_3; and calcium carbonate, $CaCO_3$, before exposure to the dye, alizarin.

Other mordants function through the formation of a complex with the dye. The complex acts as the dyeing agent. Mordant and dye in this case are exposed simultaneously to the fabric. *See* DYE; DYEING. [F.J.J.]

Morphine

An alkaloid obtained from opium, the partly dried latex from incised unripe capsules of the poppy plant, *Papaver somniferum*. The structural formula of morphine has been the object of considerable chemical research which culminated in the proposal of the full structure shown by J. M. Gulland and R. Robinson in 1925. The proposed structure was fully confirmed by synthesis by M. Gates in 1954.



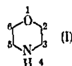
Morphine

The opium poppy is cultivated in Egypt, India, China, and Turkey, but the chief source of medicinal opium, and hence, morphine, is Asia Minor. Although opium was used before history was recorded, the first undisputed reference to poppy juice is found in the writings of Theophrastus in the third century B.C. Until well into the nineteenth century, only crude opium preparations were used medicinally. In 1803, morphine was isolated and described as the first member of a new class of substances having alkalilike, hence alkaloidal, properties.

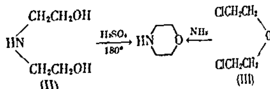
Morphine is by far the most important alkaloid of opium, and gives to the latter its predominant pharmacological characteristics. Clinically, the great usefulness of the drug lies in its depressant action resulting in a raising of the pain threshold, a sleepy feeling, a lessened perception to external stimuli, and a feeling of well-being (euphoria). The most serious objection to it is that continuous use develops addiction. In spite of this serious shortcoming, however, morphine is an excellent analgesic, and it has well-defined and indispensable uses in medicine and surgery. *See* ALKALOID. [S.M.K.]

Morpholine

A heterocyclic organic compound containing oxygen and nitrogen in a six-membered ring. Morpholine or tetrahydro-1,4-oxazine (I) is a commercially important chemical, useful as solvent, as



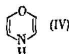
acid acceptor, and as reagent or catalyst in organic syntheses. Morpholine is a colorless liquid (bp 128.6°C) which has a characteristic odor and which is miscible with water and most common solvents. It is a secondary amine, with pK_a 8.39 at 25°C. The



ring system is constructed by cyclization of diethanolamine (II) with acid, or by cyclization of a β,β'-dichloroethyl ether (III) with ammonia. Diethanolamine (II) is obtained from ethylene oxide and ammonia. *N*-substituted morpholines are products of cyclization of *N*-substituted diethanolamines, as well as of the reaction of β,β'-dichloroethyl ethers with primary amines. When secondary amines are used, the products are quaternary morpholinium salts.

Morpholine (I) is stable to water, alkali, or acid at temperatures as high as 150°C. The ring is not cleaved by reduction. Much of the chemistry of morpholine is centered at its nitrogen atom. Monoalkylation gives *N*-alkyl derivatives; dialkylation gives quaternary salts. The molecule may be acylated and nitrosated. Morpholine adds to α,β-unsaturated carbonyl compounds, adds to ethylene oxides, and takes part in Mannich condensations. The fatty-acid salts of morpholine are emulsifying soaps that are used in floor polishes and paper coatings.

1,4-Oxazine (IV) is unknown, although some of



its derivatives are recognized. *See* HETEROCYCLIC COMPOUNDS. [W.J.GE.]

Bibliography: R. C. Elderfield (ed.), *Heterocyclic Compounds*, vol. 6, 1957.

Morphotropism

Similarity of structure, axial ratios, and angles between faces of one or more zones in crystalline substances whose formulas can be derived one from another by substitution. Examples of morphotropic substances are compounds of the type M_2SO_4, where M stands for Li, Na, K, Rb, Cs, Ag, Tl, or NH_4. Structurally, this is explained as follows. The crystal structure is essentially determined by the large SO_4 groups. The oxygen ions form a tetrahedron, with the sulfur in its center. The other ions are in cavities in between four or six

of the oxygen ions according to their dimensions. Organic compounds also show morphotropism. No definite general rules, however, can be given for its occurrence. *See* COORDINATION NUMBER; CRYSTALLOGRAPHY; CRYSTAL STRUCTURE; ISOMORPHISM (CRYSTALLOGRAPHY). [W.D.]

Bibliography: R. C. Evans, *Introduction to Crystal Chemistry*, 1939; A. E. H. Tutton, *Crystallography and Practical Crystal Measurement*, 1911.

Mortar

A plastic mixture of water and solids used to bond masonry units together. Mortar sets to a hard cementitious solid as the water reacts with the solids.

The principal solids in mortar are sand, hydrated lime (or slaked quicklime), and portland cement. The latter is absent in a pure lime mortar.

The portland cement reacts most rapidly with water, giving the mortar its initial set. Lime reacts more slowly, both with carbon dioxide (CO_2) from the atmosphere to form $CaCO_3$ and with the sand (SiO_2) to form calcium silicates.

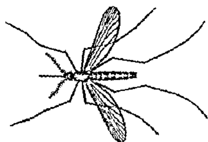
Lime is used in mortars mainly because of the good workability it imparts, leading to well-filled joints. *See* LIME (INDUSTRIAL); PORTLAND CEMENT [M.C.M.]

Moruloidea

The only class of the phylum Mesozoa. It contains two orders, the Dicyemida and Orthonectida. The class name was proposed by M. Hartmann based on the structural development of these animals. Embryonic development in the organisms proceeds as far as the morula or stereoblastula stage. Because there is a structural resemblance to the planula larva of the coelenterates, the Moruloidea are also known as the class Planuloidea. *See* DICYEMIDA; MESOZOA; ORTHONECTIDA; *see also* BLASTULATION. [C.B.C.]

Mosquito

Any member of the insect family Culicidae, order Diptera. There are about 1500 species of mosquitoes found throughout the world in warm and temperate regions; about 100 species are native to North America.



The house mosquito, *Culex pipiens*; length, female ⅛ in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

Mosquitoes are readily distinguished from related flies by their humped backs, by the presence of scales along the veins and margins of the wings and on their bodies, and by the long, slender abdomens. The aquatic larvae are commonly called wrigglers or wiggle tails.

Disease carriers. Because they are vectors for some of man's most serious diseases, mosquitoes rank high on the list of insect enemies of man. Various species are the vectors for malaria, yellow fever, dengue fever, filariasis, and some forms of elephantiasis. In addition, they create considerable annoyance by their bites and buzzing.

The various types of human malaria, a protozoan disease, are all transmitted by the bite of mosquitoes of the genus *Anopheles*. The campaign of the United Nations to eradicate malaria, one of the world's most destructive diseases, depends upon eradication of the *Anopheles* mosquito.

Both yellow fever and dengue are virus diseases transmitted by *Aedes aegypti* and other species of the same genus. Yellow fever was temporarily brought under control by a vigorous campaign against the mosquito host, but has lately been on the increase in tropical America because of the relaxation of control activities directed at this mosquito. The United States species of *Aedes* are not known to be disease carriers.

Filariasis, caused by nematodes of the genus *Wuchereria*, is transmitted by certain mosquitoes, mostly of the widespread and common genus *Culex*. Filariasis is a serious disease in tropical countries. Encephalitis is a widely distributed virus disease which is transmitted by various mosquitoes, primarily those in the genera *Culex* and *Aedes*. In addition to the various diseases of man, mosquitoes also transmit forms of malaria to birds, and they probably carry other bird and mammal diseases. *See* DENGUE FEVER; FILARIASIS; MALARIA; NEMATODA; VIRUS, YELLOW FEVER.

Life history. In a typical mosquito life history, as occurs in the genus *Culex*, groups of 100 or more eggs are laid on the surface of stagnant water. They hatch in 3–4 days into the wriggler larvae, so named because of their erratic swimming motion. Larvae undergo rapid development and may become pupae within 2 weeks or longer, depending upon temperature. The pupae are more compact than the larvae, but remain active; they transform into adults within about a week. There may be several generations each year. In cooler climates mosquitoes may overwinter as adults. In some species all the males die during the winter. However, in warm climates there is no resting period and reproduction takes place the year around.

Wrigglers hang from the surface film of the water when at rest and breathe air through a respiratory tube which protrudes from the abdomen and penetrates the surface film. Larvae of *Culex* and *Aedes* have long respiratory tubes and hang head down; those of *Anopheles* have short tubes and hang parallel to the surface.

The details of the life histories of the various species differ greatly. Larvae feed upon plant and animal plankton, adult males upon exposed plant juices and nectar, and most adult females upon vertebrate blood, which they must have before they can mature eggs. A few mosquitoes eat only plant juices and never bite animals. *See* DIPTERA; PARASITOLOGY. [J.D.B.]

Moth

Any of a large number of species of insects belonging to the order Lepidoptera, including both the butterflies and moths. The latter are distinguished by their large bodies and nocturnal habits, in contrast to the day-flying and relatively slender-bodied butterflies. The antennae of moths are varied but usually not knobbed, whereas all butterflies have knobbed antennae. Moths hold their wings in the horizontal position when at rest; the wings of butterflies are folded together over the back.

Characteristics. All of the order have sucking mouthparts (in a few species they are vestigial) as adults, and chewing mouthparts as larvae. The larvae are called caterpillars and almost all are plant eaters. All Lepidoptera have complete metamorphosis, and the caterpillar passes through several molts before entering the resting, or pupal, stage. In contrast to butterflies, most moths pupate in a cocoon, and their pupae are frequently called chrysalids

Although some of the largest and showiest of our insects are moths, many moths are well-camouflaged, drab insects that frequently escape notice. Moths usually have but one generation a year, and overwinter in the pupal or larva stage. However, some spend the winter as adults, and a few others as eggs.

Best known of the showy moths are the giant silkworm moths of the family Saturniidae. Included here are the luna, cecropia, io, polyphemus, and promethea moths. The family Sphingidae, the sphinx or hawk moths, include a large number of narrow-winged moths which are frequently gray above but show brilliant red, black, yellow, and orange markings in flight. This family also includes the [illegible] sphinx moth, *Protoparce quinquemaculata*. The larvae of this family are often called hornworms because of the prominent spines borne on the eighth abdominal segment of most species.

Pests. The large family Noctuidae (Phalaenidae) includes some of the most destructive agricultural pests, notably the army worm, cutworm, corn earworm, and cabbage looper. The family Liparidae includes several serious pests of forest and shade trees, such as the white-marked tussock moth and two introduced species, the gypsy moth and brown-tailed moth, both of which have caused extensive damage to trees in the northeastern United States. Control measures undertaken against the last two have been extremely expensive and quite destructive to wildlife.

Other well-known moth pests include the clothes moth, tent caterpillar, cankerworm, and the European corn borer. *See* CANKERWORM; CATERPILLAR; CORN.

Perhaps the only species helpful to man in the entire group is the silkworm moth, *Bombyx mori*, the only species of the family Bombycidae. *See* ARMY WORM; BUTTERFLY; CUTWORM; INSECTA; LEPIDOPTERA; SILK. [J.D.B.]

Motion

If the position of a material system as measured by a particular observer changes with respect to time, that system is said to be in motion with respect to the observer. Absolute motion, then, has no significance, and only relative motion may be defined; what one observer measures to be at rest, another observer in a different frame of reference may regard as being in motion. *See* FRAME OF REFERENCE; RELATIVE MOTION.

The time derivatives of the various coordinates used to specify the system may be used to prescribe the motion at any instant of time. How the motion develops in subsequent instants is then determined by the laws of motion. In classical dynamics it is supposed that in principle the motion and configuration of the system may be specified to an arbitrary precision, although in quantum mechanics it is recognized that the measurement of the one disturbs the other.

For a system of f degrees of freedom, the motion may be represented by a point in an f dimensional velocity space, the coordinates of which are the time rates of change of the coordinates that describe the configuration of the system. For a system under no forces that is described by rectangular cartesian coordinates, these time derivatives are constants. For a single particle, this result is the first of Newton's laws of motion, namely that a particle remains at rest or in a state of uniform motion unless acted upon by an external force.

The most general theory of motion that has yet been developed is quantum field theory, which combines both quantum mechanics and relativity theory, as well as the experimentally observed fact that elementary particles can be created and annihilated. *See* QUANTUM FIELD THEORY; *see also* DEGREE OF FREEDOM (MECHANICS); DYNAMICS; EULER'S EQUATIONS OF MOTION; HAMILTON'S EQUATIONS OF MOTION; HARMONIC MOTION; KINEMATICS; KINETICS (CLASSICAL MECHANICS); LAGRANGE'S EQUATIONS; MOTION, RECTILINEAR; NEWTON'S LAWS OF MOTION; OSCILLATION; PERIODIC MOTION; QUANTUM MECHANICS; RELATIVITY; ROTATIONAL MOTION. [H.C.C.]

Motion, rectilinear

Motion is defined as continuous change of position of a body (*see* MOTION). If the body so moves that every particle of the body follows a straight-line

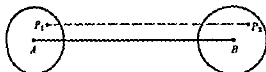


Fig. 1. Rectilinear motion. All points move parallel to the center of mass.

path, then the motion of the body is said to be rectilinear.

When a body moves from one position to another, one may describe the effect in terms of motion of the center of mass of the body from a point A to a point B (see Fig. 1). If the center of mass of the body moves along a straight line connecting the points A and B, then the motion of the center of mass of the body is rectilinear. If the body as a whole does not rotate while it is moving, then the path of every particle of which the body is composed is a straight line parallel to or coinciding with the path of the center of mass, and the body as a whole executes rectilinear motion. This is shown by the straight line connecting points P_1 and P_2 in the figure. *See* CENTER OF MASS.

Rectilinear motion is an idealized form of motion which rarely, if ever, occurs in actual experience, but it is the simplest imaginable type of motion and it thus forms the basis for the analysis of more complicated motions. However, many actual motions are approximately rectilinear and may be treated as such without appreciable error. For example, a ball thrown directly upward may follow, for all practical purposes, a straight-line path. The motion of a high-speed rifle bullet fired horizontally may be essentially rectilinear over a short length of path, even though in its larger aspects the ideal path is a parabola (*see* BALLISTICS, EXTERIOR). The motion of an automobile traveling over a straight section of roadway is essentially rectilinear if minor variations of path are neglected. The motion of a single wheel of the car is not rectilinear, although the motion of the center of mass of the wheel may be essentially so.

Curvilinear motion. Rectilinear motion must be distinguished from curvilinear motion. In the former the direction of motion is constant, while in the latter the direction of motion is continuously changing. If a body is moved from A to B along a curved path (the solid line of Fig. 2) the resulting position of the object is the same as if the body had undergone rectilinear motion from A to B (dotted line) followed by a rotation. Thus any displacement of a body may always be described in terms of rectilinear motion plus a rotation. *See* ROTATIONAL MOTION

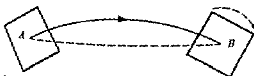


Fig. 2. Curvilinear motion. It is equivalent to rectilinear motion plus a rotation.

Motion with constant velocity. When a body moves from one location to another it is said to be displaced. The linear displacement is the distance from the first position to the second position. *See* DISPLACEMENT (MECHANICS). Motion cannot be instantaneous; it must involve a time interval during which the displacement takes place. This interval is called the elapsed time. When the ratio of the displacement to the elapsed time is constant regardless of how small an interval is chosen, the motion is uniform and the body is said to move with a constant velocity since neither the magnitude of the motion (its speed) nor the direction of the motion is changing. Motion with constant velocity is the simplest form of rectilinear motion.

Average velocity; acceleration. Rectilinear motion in general may be motion with an increasing velocity (positive acceleration) or with a decreasing velocity (negative acceleration or deceleration) or with a variable velocity (*see* ACCELERATION; VELOCITY). When the velocity is not constant, the ratio of the displacement to the elapsed time is the average velocity for the interval. The limiting value of the ratio of the displacement to elapsed time as the interval is made smaller and smaller is the instantaneous velocity, sometimes called the velocity at a point.

The displacement s of a moving body in a given time interval t is the average velocity v_{av} multiplied by the time interval, or in symbols,

$$s = v_{av}t$$

If the velocity is uniformly increasing or decreasing, the average velocity is half the sum of the initial velocity v_0 at the beginning of the interval and the final velocity v_f at the end of the interval

$$v_{av} = \frac{v_0 + v_f}{2}$$

The gain (or loss) of velocity during the interval is $v_f - v_0$ and the gain or loss per unit time is called the acceleration a and is

$$a = \frac{v_f - v_0}{t}$$

From this, the gain (or loss) of velocity $v_f - v_0$ is equal to the acceleration times the elapsed time.

$$v_f - v_0 = at$$

For a body uniformly accelerated from rest, the initial velocity v_0 is zero and the average velocity is $v_f/2$; this equals $at/2$ since the final velocity of a uniformly accelerated body starting from rest is $v_f = at$. Consequently, the displacement or space traversed by a uniformly accelerated body starting from rest is

$$s = v_{av}t = \left(\frac{v_f}{2}\right)t = \left(\frac{at}{2}\right)t = \tfrac{1}{2}at^2$$

If the initial velocity of the body is not zero, the displacement due to the continuing action of the initial velocity must be added to that due to the acceleration:

$$s = v_0 t + \frac{1}{2}at^2$$

The units in which velocity is usually measured are meters per second (m/sec), feet per second (ft/sec), or miles per hour (mph). For acceleration, the most common units are meters per second per second (m/sec²) or feet per second per second (ft/sec²).

The foregoing formula is the solution to the basic problem in the kinematics of rectilinear motion. The problem is as follows: given the initial conditions, that is, the initial velocity v_0 and the acceleration a, find the position of the body after any elapsed time t. This requires that both the velocity and acceleration be vector quantities denoting direction as well as magnitude and that the initial position of the object be known. If this initial position is denoted by s_0, then in general the position s after any elapsed time t is

$$s = s_0 + v_0 t + \frac{1}{2}at^2$$

For rising and falling bodies, it is necessary only to substitute for a the known acceleration of gravity g, which is approximately 9.8 m/sec² or 32.2 ft/sec² for bodies near the surface of the earth.

Newton's laws. When all the forces acting on a body are balanced, there is no net unbalanced force, and the body either remains at rest or continues to move in the same direction with constant velocity (Newton's first law of motion; *see* NEWTON'S LAWS OF MOTION). Whether at rest or in motion under the action of balanced forces, the body is said to be in equilibrium. When an unbalanced force of constant magnitude and direction acts upon a body, the body will be continuously accelerated with a constant acceleration (Newton's second law of motion). A freely falling body near the surface of the earth would be an almost ideal example if it were not for the resisting force of the air (*see* FREE FALL). Since this resistance increases with velocity, a point is reached sooner or later where the resisting force exactly balances the downward pull of gravity. The body is then in equilibrium, the acceleration is zero, and the body continues to fall with a constant velocity called the terminal velocity. Other more complicated motions can be described in terms of more complicated force functions, such as one in which the force (1) varies in direct proportion to the distance the body has traveled, (2) varies as the square of the distance, or (3) varies inversely as the distance, or (4) varies inversely as the square of the distance.

Force and motion. The basic formula relating force and motion and expressing both of Newton's first two laws of motion is $F = ma$ or $a = F/m$; it indicates that the acceleration produced in any body of mass m by an unbalanced force F is directly proportional to the force and inversely proportional to the mass. If the force is zero, the acceleration is zero, and the motion (if any) is nonaccelerated; that is, has constant velocity. In the symbols of calculus

$$F = m\frac{d^2s}{dt^2}$$

If the appropriate force function is put into this [illegible]

See FORCE. [R.D RU.]

Motivation

The intentions, desires, goals, incentives, interests, values, beliefs, and attitudes, that determine human and animal behavior. An inquiry about a man's motives seeks an answer to the question "Why?" The answer has important consequences. If, for example, a man kills another with intent, he may go to the electric chair; if he carelessly kills a man, he may be convicted of manslaughter and be given a lighter sentence. The difference is one of motivation.

Within technical psychology there are broad and narrow definitions of motivation. Broadly conceived, the study of motivation is a search for the determinants of behavior. From this point of view all behavior is motivated in the sense that it is causally determined. Some psychologists, however, distinguish between motivated and unmotivated behavior, thus implying a narrower definition. The restricted definition usually describes motivated behavior as persistent, goal-oriented activity. From this point of view random, aimless, reflexive, and some abnormal activities, are regarded as lacking motivation.

Broadly viewed, motivation may be defined as the process of arousing, sustaining, and regulating behavior. This process has at least three important aspects. First, motivation is an energizing process, which means that behavior is aroused by the release of energy within the tissues. Second, motivation is a process of regulating and directing behavior. This aspect is apparent in all activity but especially so in persistent purposive behavior. Third, the process of motivation has a developmental aspect. It is necessary to consider the growth of motives and other determinants.

The study of motivation can be carried on within the framework of physical, biological, psychological, or social science. This makes the study complex and may explain why not all psychologists have agreed upon the basic definitions of motive, drive, set, and other dynamic concepts.

DYNAMICS OF BEHAVIOR

The term dynamic psychology commonly designates a psychology of human motivation, including both normal and abnormal, that stems from the work of Freud and psychoanalysis. The following terms indicate the nature and scope of this kind of

psychology: unconscious motivation, conflict, repression, projection, sublimation, fixation, identification, rationalization, and compensation. That these terms are commonplace is witness to the profound influence of Freud upon the psychology of human motivation.

The concept of unconscious conflict, for example, is important in explaining such abnormal phenomena as amnesia, functional paralysis, phobia, *multiple* personality, and the like. The concept of *repression* helps to explain the causation of compulsions, obsessions, and phobias.

Frustration builds up tension, with diverse effects upon behavior (*see* AGGRESSION; FRUSTRATION; NEUROSIS).

Principle of homeostasis. A physiological basis for the study of needs and drives is offered by the principle of homeostasis. Claude Bernard, in 1859, pointed out that the cells of the body live within an internal environment, or milieu interiéur, consisting of the body fluids, notably blood and lymph. An essential condition for survival of the cells, and hence the total organism, is that the milieu interiéur maintain specific physicochemical properties. W. B. Cannon coined the term homeostasis to designate a continuing process which attempts to maintain a relatively stable state of the body fluids. Within the blood, for example, there is a homeostasis of temperature, of blood sugar, water content, oxygen content, and pH. These relatively stable states are maintained through cooperative action of brain, nerves, heart, lungs, skin, kidneys, spleen, and other organs.

C. P. Richter showed that the behavior of man and animal is, in many ways, directed toward the maintaining of homeostasis. For example, if the pituitary and thyroid glands of rats are surgically removed, the rats lose their ability to produce adequate amounts of body heat. They make up for this deficiency by building nests to insulate themselves against heat loss. Richter argued that the effort to maintain a stable internal environment constitutes one of the most universal and powerful of behavioral urges. The basic drives, to obtain air, water, food, rest, and activity, to urinate and defecate, to avoid extremes of heat and cold, are clearly directed toward the maintaining of homeostasis (*see* HOMEOSTASIS).

Needs. The concept of needs can be defined by reference to the *principle of homeostasis.* In order to maintain homeostasis an organism must obtain from the milieu extérieur, or external environment, definite quantities of oxygen, water, fat, protein, carbohydrate, minerals, and vitamins. Further, for warm blooded species the internal body temperature must be maintained within a relatively narrow range despite wide variations in external temperature. Waste products of metabolism must be eliminated. Sleep, rest, and opportunity for activity must be possible. A homeostatic need is something the organism requires to maintain a stable physicochemical state within the body fluids.

H. A. Murray distinguished between viscerogenic, or primary, and psychogenic, or secondary, needs. Viscerogenic needs originate in tissue conditions; psychogenic needs depend upon the social environment. Examples of psychogenic needs are those for achievement, recognition, dominance, aggression, affiliation, love, and security.

For Murray, as for some other psychologists, a need is dynamic in the sense it motivates behavior. A need, however, may be either a dynamic or nondynamic concept. A dietary deficiency can affect growth or reproduction without directly influencing behavior. A deficiency of vitamin E, for example, may impair reproduction in laboratory animals without affecting their behavior. C. P. Richter has shown that rats on certain vitamin-deficient diets select foods which remedy the specific deficiency; but this selectivity does not hold true for all vitamins and minerals.

Organic drives. Persisting bodily states, such as hunger, thirst, or sexual excitement, that influence behavior and conscious experience are called organic drives. A drive leads to the development of persistent goal-oriented behavior that continues until a consummatory response has been made that reduces or removes the motivation For example, a thirsty man, lost on a desert, continues his search for water until he finds it and drinks. In human beings the drive is experienced as an appetite, or craving, with both cognitive and cathectic aspects.

E. C. Tolman, following W. Craig, classified drives into two groups on the basis of behavior,

response is not made. An aversion is also a state of agitation but one that continues as long as the aversive stimulus object is present. Appetites are cyclic; they depend upon recurring organic conditions; they lead to the development of approach patterns; they are terminated by appropriate goal-responses. Aversions, contrastingly, depend upon external conditions that arise from time to time; they are manifested by development of avoidance patterns; they are terminated by acts instrumental in removing or avoiding noxious stimulation. According to Tolman, two main human aversions are fright and pugnacity.

R. S. Woodworth, in 1918, introduced the term drive to American psychology. He distinguished between drive and mechanism. Drive is the physical energy that makes a machine go; mechanism is a regulative structure.

When psychologists speak of drives, in the plural sense, however, they are thinking not in terms of physical energy but in terms of organic mechanisms that both energize and regulate behavior. Specific organic drives can be distinguished in terms of their bodily mechanisms, as well as in terms of behavior. Examples of primary organic drives are hunger for food, thirst, air hunger, the drive to maintain an optimal temperature, the drives

urinate and to defecate, the drive to rest when fatigued and to be active after prolonged inactivity, the drive to sleep, the drive to copulate, the female drive to nurse the young. These drives, revealed in behavior, have a physiological counterpart which can be described in physicochemical and anatomical terms.

Behavior theorists regard drive as a hypothetical construct rather than as a concrete bodily state. For example, C. L. Hull postulates a drive stimulus (S_d) that is derived from a need state and which has properties similar to other stimuli. The drive (D) itself has both energizing and reinforcing properties and is thus importantly related to learning and performance. B. F. Skinner, however, is willing to dispense with the concept of drive, describing behavior in terms of present and antecedent conditions (*see* HUNGER; PAIN, CUTANEOUS; PAIN, DEEP; REPRODUCTIVE BEHAVIOR; SLEEP).

Incentive motivation. The behavior of organisms is an interlocking relation between organic and environmental conditions. It is misleading to separate sharply the two realms since behavior is governed by environmental as much as by organic conditions.

By definition, an incentive is an environmental motivation. It may be a goal object such as food in the presence of a hungry animal, water near a thirsty animal, or a receptive mate. The incentive may also be a form of stimulation, such as spurs forced into the flank of a horse, or an electric shock to the feet of a laboratory rat.

Experiments have shown that human performance depends upon a variety of incentives and motives. Among the factors which have been studied are the following: having or not having a goal, understanding the task, interest, eagerness to succeed, and effort. Social incentives include prizes and decorations, praise and reproof, encouragement and discouragement, rewards and punishments, competition, rivalry, cooperation, various social recognitions, and others. In addition to these motivating factors, physical conditions, both environmental and organic, affect the level of human performance. Examples are temperature, humidity, air movement, illumination, distraction, music, fatigue, boredom, sleepiness, illness, and drugs.

Laboratory experiments with animals have dealt with a variety of incentive conditions, including the quantity of food offered to a hungry animal as a reward; the palatability of the food incentive; the interval of delay between a performance of a learned task and the reward; the relevance of incentive to drive state; and the symbolic reward, for example, a reward of poker chips. Chimpanzees can be adequately motivated to work for poker chips provided these chips can be deposited in a slot machine that delivers food. Electric shocks are widely employed as a negative incentive.

If a rat is placed in an apparatus with a grid on the floor and given a shock through his feet, there is an immediate emotional response. The animal may jump, squeal, urinate, and defecate; its heart rate and blood pressure may rise; and the rat may move about actively or "freeze" and tremble (fear). If placed on the apparatus a day later, this time without shock, the environmental stimulus cues evoke a learned emotional response (anxiety). A rat, so motivated, may learn an instrumental act, like pressing a lever, that permits him to escape from a threatening compartment into a safe one. N. E. Miller and O. H. Mowrer believe that the an- [illegible]

forces acts that are instrumental in reducing the basic motivation.

The stimulus-response principle. If behavior is viewed as a physical process, then the factors that arouse, sustain, and regulate behavior must also be regarded as physical. From the physical point of view, behavior is a form of movement and the problem of motivation is that of describing how energy is released in the tissues and how behavior is regulated so that it becomes goal-oriented.

When viewed in the light of physical science, every stimulus that elicits a response appears to be motivating. In a simple reflex, the stimulus releases energy and the response is regulated by the reflex arc and other bodily structures. In persistent [illegible] is similar in [illegible], however, [illegible] is depriva- vary in intensity with [illegible] tion and satiation. In postural sets and adjustments an important source of stimulation is found in the movement of tensing of muscles, tendons, and joints. In these and other forms of behavior, motivation is equated with stimulation.

The stimulus-response theory of motivation must be accepted with caution if stimuli are thought of as always coming from the external environment. According to an older view, the organism is a passive structure that is activated by stimuli. Modern studies in electroencephalography, however, have shown that brain cells are continuously active, even during sleep, and that brain activity can initiate movement apart from external stimulation. Further, the stimulus-response theory must take account of both chemical sensitization and central emotional processes.

Chemical motivation. It is well known that hormones affect both growth and behavior. For example, if the testes and ovaries of young chicks are transplanted, the animals bearing the testes, regardless of original sex, develop physical traits of the male and masculine patterns of behavior; the animals bearing the ovaries develop a female form and feminine behavior. The gonadal hormones thus determine sexual behavior as well as physical development.

Chemical determinants are of primary importance in food hunger, satiation, air hunger, fatigue, anger, fear, sexual and maternal activity, and other organic drives.

Chemical motivation appears during embryonic development prior to response to external stimulation. G. E. Coghill, for example, demonstrated that the embryonic toadfish responds to a change in concentration of CO_2 in the surrounding water by rhythmic, integrated, muscular contractions. These changes in behavior can be demonstrated well before sensory structures have sufficiently matured to respond to external stimulations. Coghill demonstrated, at least for one species, that internal chemical motivation ontogenetically precedes motivation mediated by the nervous system.

It should be added that research in the growing field of psychopharmacology has demonstrated repeatedly the great importance of chemical determinants of behavior and of mood (*see* PSYCHOPHARMACOLOGIC DRUGS). It is common knowledge that a morphine addict develops a powerful appetite for the drug. Such an addiction differs from a simple habit in that there is chemical motivation as well as neural.

Psychological hedonism. Hedonism is an ancient doctrine stating that activities felt as pleasant tend to be continued and repeated; those felt as unpleasant tend to be discontinued and avoided. In its traditional form the hedonic theory of motivation implies that conscious feelings influence the course of human action. Although animals act as if they experience distress and occasionally delight, it is impossible to know for certain just what they feel and even that they are conscious in the sense that adult humans are.

The hypothesis that hedonic processes have an objective existence and that they influence behavior, however, has been made plausible by discoveries of a neural basis for reward and punishment. J. Olds discovered that if certain points in the subcortical region of the rat's brain are stimulated through continuously implanted electrodes, the animal acts as though he were being rewarded. A rat will press a bar repeatedly, if this action closes a circuit that stimulates a reward center. If the experimenter breaks the circuit so that internal stimulation ceases, the bar-pressing activity gradually extinguishes. In this work electrodes are permanently implanted within the brain and stimulation is provided through a light, flexible cord suspended from the ceiling of the apparatus. Reward centers are located in subcortical regions, especially in the septal area, the amygdaloid complex, the hypothalamus, and in nearby regions (*see* NERVOUS SYSTEM).

In a series of experiments, J. M. R. Delgado, W. W. Roberts, and N. E. Miller found that electrical stimulation, through implanted electrodes, of certain points within the brain of the cat, appears to act like punishment. For example, cats learned to rotate a wheel at the sound of a buzzer when this action broke a circuit and discontinued stimulation through the implanted electrodes. These and other investigators have also reported the existence of neural centers which, when stimulated, are neither rewarding nor punishing. Further, there is some evidence that continuous stimulation of certain points may be alternately rewarding and punishing. More research is needed to clarify the picture.

Apart from this physiological work, P. T. Young, in studies on rats concerning the development of food preferences, feeding habits and appetites, postulated the objective existence of affective processes. These are processes formerly known only as subjective feelings of pleasantness or unpleasantness. Some of the principles disclosed by experiments based upon the postulate of objective hedonic processes are these:

1. Affective arousals are both positive and negative. If a naïve animal develops an approach or maintaining pattern of behavior, the existence of a positive affective arousal may be inferred. If a naïve animal develops a pattern of avoidance or escape a negative arousal may be inferred. If neither approach nor avoidance develops, affective neutrality may be inferred.

2. The strength of a recently organized approach motive depends upon the frequency, intensity, duration, and recency of positive affective arousals. This is shown by the fact that the speed of locomotion with which a rat approaches a sugar solution varies with the concentration of solution, the frequency of contact with it, the duration of contact, and the recency of contact. A similar relation is assumed to exist between the strength of newly organized avoidance motives and the intensity of painful distress.

3. If animals are given a choice between two foods, the preference that develops reveals the relative palatability of those foods. The preferred food has a higher hedonic value than the nonpreferred. Incidentally, preference tests based upon the quantity of food ingested are ambiguous because the quantity ingested depends not only upon palatability but also upon internal changes as satiation is approached, upon feeding habits, and other factors.

Activation and affectivity. As a result of researches in electroencephalography, D. B. Lindsley formulated an activation theory of emotion. The theory actually takes account of much more than emotion, for it is concerned with sleep, coma, and pathological conditions, as well as with variations in normal alertness.

It is enough to note that Lindsley postulated different degrees of activation within the reticular system. Anatomically, this system is a network of nerve fibers in subcortical regions. The neural impulses from visceral and somatic receptors may

into the reticular formation, along collaterals, aff
the cortex through the diffuse projection s

Cortical excitation via the reticular activating system alerts the cerebral mechanism but conveys no specific information. Such excitation produces a generalized pattern of activation in the cortex.

In terms of cortical activity, as revealed by the electroencephalogram, a high level of activation is characterized by desynchronization and inhibition of the alpha (brain wave) rhythm. Lower levels of activation are characterized by increased synchrony and rhythmic alpha waves. The very low levels of lethargy and sleep are characterized by large, slow waves and by relatively high synchronization.

The different levels of activation can be represented as points upon a vertical continuum (Fig. 1). The lowest levels correspond to degrees of general activity observed in coma and sleep. Higher levels correspond to the range of normal alertness. The highest levels correspond to degrees of activity presented in great excitement, rage, and other emergency emotions.

A single continuum of activation does not represent the difference between positive and negative affective arousals. A second continuum, therefore, is needed (Fig. 2). The illustration represents the fact that there are positive and negative forms of affectivity that range from intense negative affectivity, or unpleasantness, through indifference to intense positive affectivity, or pleasantness. Arrows indicate two opposed directions of hedonic change: change toward the positive end of the continuum and change toward the negative end. Actually there are four kinds of hedonic change that are psychologically important: increasing and decreasing positive affectivity; increasing and decreasing negative affectivity. These illustrations do not represent the many specific patterns of response that are evoked during reflexive, instinctive, and emotional behavior (*see* ELECTROENCEPHALOGRAPHY; EMOTION).

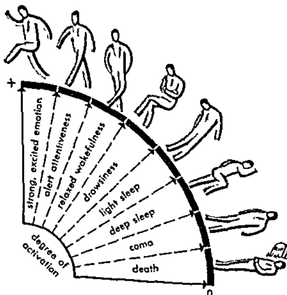


Fig. 1. The continuum of activation.



Fig 2. The hedonic continuum.

REGULATION AND DIRECTION

Another very important aspect of motivation is the regulation and direction of behavior.

Set. The concept of set refers to an orientation or inner disposition of the organism. A bodily set may be directly observed. A runner, for example, toes the line, maintains his stance, awaits the gun; his bodily posture is a neuromuscular set involving observable changes in muscle tonus as well as a neural readiness. A purely neural set may persist or at least be repeatedly evoked or reinstated in the absence of observable muscular tension. For example, if a man has determined to study engineering, his determination persists as a regulative and controlling factor even when all muscles are completely relaxed in dreamless sleep.

The concept of set is important in relation to attention and will. Involuntary attention is determ... moving ob-

a set. Voluntary attention is accompanied by a sense of effort as in attentively steering a car into the glare of an oncoming automobile or in persistently studying a lesson despite distractions. Again, a will to act is a kind of set or determination. A strong-willed man does not easily change his determination once he has made up his mind. A weak-willed person is indecisive and vacillating; he has difficulty in reaching a decision and holding to a resolution.

Within technical psychology the term set has ... three stand out as ... s a set to ... aiting the arrival of a friend he is set to perceive a familiar face. (2) Intention is a set to act. After a child has been instructed to carry out an errand he is predisposed, or set, to carry out the specific instruction. (3) There is a central set. A problem or a topic of conversation regulates the interplay of associations.

In general, it can be said that a set regulates and directs activity. It acts to sustain, organize, accentuate, and select.

Retroflex action and feedback. In his book upon human motivation, L. T. Troland emphasized retroflex action. This process can be illustrated by a familiar example. If a child sees a candle flame, he is likely to reach for it as for any other bright object; if he gets burned, there is a nociceptive

(painful) feedback which modifies neural conductance in the brain so that on future occasions the reaching response is checked or modified. The general principle of retroflex action is this: Whenever an organism makes a response, this produces new stimuli and, frequently, an affective arousal.

The affective arousal is an important component of retroflex action. Thus if a boy eats some green apples, his behavior is followed by acute distress that leaves his nervous system less ready to allow him to eat green apples again. If he tastes honey, the delightful flavor makes him more ready to try it at the next opportunity.

The regulatory principle in these instances is clearly hedonic. Troland believes that retroflex action is of basic importance in learning. It is something more than an empirical law of effect.

Although the effective arousal is an important component of retroflex action, the sensory feedback is also important in the regulation and control of behavior. Examples of regulation through sensory feedback are numerous. A few will be considered. A dog's path is determined by the behavior of a rabbit he is chasing. If the rabbit turns to the right, the dog turns right; if the rabbit dashes under a fence, the dog jumps over the fence; if the rabbit goes into a hole, the dog approaches and starts to dig. Thus the dog's behavior is regulated by information fed into the nervous system from the visual world.

Regulation by auditory feedback can be illustrated by the flight of bats. These animals are nearly blind. They forage for food at night, traveling many miles and then returning at dawn to their caves or sleeping places. Bats emit exceedingly rapid bursts of sounds with ultrasonic frequencies of 30,000–50,000 cps. These high frequency sounds are reflected to the ears of the animals by solid objects lying in their flight path. By responding to the reflected waves, bats avoid collisions with the limbs of trees and other obstacles; this represents a natu- [illegible]

such as reciting the alphabet or performing a dance, one response furnishes a stimulus pattern that arouses the next response. In the absence of proprioceptive feedback the learned series of responses could not occur.

Proprioceptive feedback has an activating as well as a regulative function. All sensory stimuli, as G. L. Freeman has pointed out, affect the basic level of activation, or tension. Every individual has a characteristic level of tension; this level is raised by external and internal stimuli and lowered by reduction of such stimuli. Every skilled act has an optimal level of tension and one can speak meaningfully of undermotivation, overmotivation, and optimal motivation. The tension level, incidentally, varies with a good many factors, including age, metabolic rate, postural set, frustration, and distraction, as well as with external stimulation and proprioceptive feedback.

Sociocultural regulation. The basic human needs in hunger, thirst, sex, defense, aggression, curiosity, play, and the like, are typically met within a society. The biological and social determinants of behavior are so closely interrelated that it is difficult and perhaps useless to try to separate them. In every society there are rules that regulate the production, preparation and eating of food; rules that regulate sexual conduct and care of the young; rules of etiquette; rules for religious behavior; rules that regulate style of dress; and so on.

Behavior is regulated by the cultural pattern and by the actions of other persons in the group. The cultural pattern is made up of elements such as a language, a system of beliefs about the natural and supernatural worlds, specific techniques for healing the sick, educating the young, fighting, and so forth. These elements can be abstracted logically from concrete social reality. The elements of culture are internalized by individuals growing up [illegible]

of social motivation must take account of interpersonal orientations. An individual has an orientation toward other persons, toward himself, and toward concrete objects. An orientation has two main aspects, cognitive and cathectic. The cognitive aspect relates to what we perceive, know, believe, and imagine. The cathectic aspect relates to motivation, to value, interest, and action. An orientation, therefore, includes not only knowledge but a readiness to do something with respect to the object. Various other terms refer to a person's orientation, such as attitude, value, interest, loyalty, and motive.

Attitudes toward one's self are of tremendous psychological importance. A person may develop an attitude of inferiority toward himself. Alfred Adler pointed out that the attempt to compensate for a feeling of inferiority may be a deep-seated and persistent motivation. An attitude of self-confidence makes for efficiency of work.

When a person regards a task or an object as his own, he is said to be ego-involved. Ego-involvement is an important element in social and personal motivation.

Cognition and motivation. D. O. Hebb distinguished between cue and arousal functions. Cue function refers to association, to symbolic meaning, to something cognitive. Arousal function refers to drive. Drive, he stated, is an energizer, not a guide; an engine, not a steering gear; a propeller, not a rudder. Drive gives impetus to behavior, but drive is always related to a cue function which is directive and regulative. Cue function is apparent in motor control and in cognitive performances. Thus drive is energizing; cue function is regulative and directive.

Studies of the relation between cognition, or cue function, and motivation, or arousal function, have been made chiefly in the areas of perception and fantasy.

Experimenters have studied how needs, anxieties, and organic states influence perception. For example, J. C. Gilchrist and L. S. Nesberg required subjects to match the visual brightness of food objects and drinks that were projected successively on a screen. They found that the subjects unknowingly adjusted the apparatus to make the pictures of need-related objects brighter than the pictures of objects that were not related to their need states. E. McGinnies exposed socially taboo and nontaboo words for intervals too brief for adequate recognition. At the same time he recorded the galvanic skin response as an index of affectivity. The data showed that taboo words evoked an affective response with exposure times too brief for complete recognition and verbal report. The resistance to recognizing and reporting taboo words McGinnies called perceptual defense.

The investigations of D. C. McClelland, J. W. Atkinson, and others, have grown out of Freud's claim that fantasy reveals basic, often unconscious, human motivations. In fantasy a subject is not bound by perceptual reality. He is free to express his wishes, anxieties, feelings, beliefs, attitudes, needs, and values.

The interpretation of ink blots, pictures, and other ambiguous material yields a kind of data that is uniquely human. The utilization of these data is something of an art, requiring special training, despite the claims of devotees that scoring of projec- [illegible] tion test (TAT). In this test the subject is instructed to interpret a set of standard pictures. His interpretation reveals, unwittingly, information concerning his personal motives and emotions. With the TAT technique, studies have been made upon the influence of hunger, the achievement motive, the motives for power, sex, fear, and aggression.

MOTIVATION AND LEARNING

An adequate account of motivation must consider the developmental aspect. At a given stage of development, the determinants of behavior depend upon a structural organization that has grown through the processes of maturation, exercise, and learning, and that may have changed through traumata. Instinctive behavior, unconditioned reflexes, and tropisms are unlearned forms of behavior. They are motivated, *that is*, casually determined, by present and antecedent conditions.

Learning is a growth process that must be studied in relation to motivation. Psychologists agree that learning *occurs only* when an organism responds although responses need not involve overt movement. A completely passive organism learns nothing. If it is assumed that learning occurs only during activity and that all behavior is motivated, then it follows that learning does not occur apart from motivation. Of course, in a narrower view, it can be shown that unmotivated learning occurs. It is obvious that instrumental acts and persistent goal-oriented patterns of behavior are learned.

Externalization of drive. E. E. Anderson argued that as behavior develops there is a gradual increase in environmental control. In early stages of development, behavior is largely determined by organic drives. In later stages environmental cues arouse learned patterns even in the absence of drives. This increasing environmental control he called externalization of drive. Externalization spreads from one stimulus object to another until finally a great many stimulus objects arouse the learned response. Thus hunger is an organic drive, but the instrumental activities of feeding are learned and increasingly controlled by environmental cues. If externalization were complete, the behavior which leads to feeding could be aroused in full strength by presenting the appropriate environmental cues in the absence of organic hunger.

The doctrine of externalization of drive recognizes the importance of environmental motivations and that learned acts may be functionally independent of their original internal motivations. G. W. Allport has argued that the motives of a human adult are functionally autonomous; that they are independent of organic drives even though such drives may have played a part in their development. He writes that as a man advances from infancy to maturity the character of his motivation alters so radically that one may speak of adult motives as having supplanted those of infancy.

It may be impossible, as Allport contends, to trace adult motives back to organic drives. There are, however, other sources of motivation than organic drives that must be considered in a developmental account of motivation. The sociocultural determinants, the hedonic determinants, the rôle of cognition, and other factors must be considered in any comprehensive account of the growth of human motives. *See* PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL. [P.T.Y.]

Bibliography: J. W. Atkinson, *Motives in Fantasy, Action, and Society*, 1958; M. R. Jones (ed.), *Nebraska Symposium on Motivation*, 1953—; G. Lindzey (ed.), *Handbook of Social Psychology*, vol. 2, 1954; D. C. McClelland, *Studies in Motivation*, 1955; C. P. Stone (ed.), *Comparative Psychology*, 3d ed., 1951; R. S. Woodworth, *Dynamics of Behavior*, 1958; P. T. Young, *Motivation of Behavior*, 1936.

Motor, electric

An electric rotating machine which converts electric energy into mechanical energy. Because of its many advantages, the electric motor has largely replaced other motive power in industry, transportation, mines, business, farms and homes. Elec-

tric motors are convenient, economical to operate, inexpensive to purchase, safe, free from smoke and odor, and comparatively quiet. They can meet a wide range of service requirements—starting, accelerating, running, braking, holding, and stopping a load. They are available in sizes from a small fraction of a horsepower to many thousands of horsepower and in a wide range of speeds. The speed may be fixed (or synchronous), constant for given load conditions, adjustable, or variable. Many are self-starting and reversible. For uniformity and interchangeability, motors are standardized in sizes, types, and speeds. See ELECTRIC ROTATING MACHINERY.

Electric motors may be alternating-current (ac) or direct-current (dc). There are many types of each. Although ac motors are more common, dc motors are unexcelled for applications requiring simple, inexpensive speed control or sustained high torque under low-voltage conditions.

Motor classification. Motors are classified in many ways. The following classifications show some of the many available variations in types of motors.

1. Size: fractional or integral horsepower.

2. Application: general purpose, definite purpose, special purpose, or part-winding start. May be further classified as crane, elevator, pump, etc.

3. Electrical type: alternating-current induction, synchronous, or series; direct-current series, shunt or compound.

4. Mechanical protection and cooling: (a) Open: dripproof, splashproof, semiguarded, fully guarded, externally ventilated, pipe ventilated, weather protected. (b) Totally enclosed: nonventilated, fan cooled, explosionproof, dustproof, ignitionproof, waterproof, water cooled, water-air cooled, air-to-air cooled, pipe ventilated, fan cooled guarded.

5. Speed variability: constant speed, varying speed, adjustable speed, adjustable varying speed, multispeed.

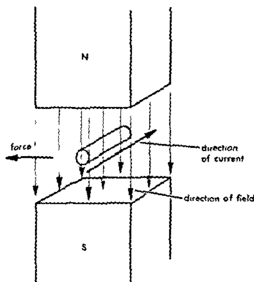
6. Mounting: floor, wall, ceiling, face, flange, vertical shaft.

Motor characteristics. Each electrical type of motor has its own individual characteristics. Each motor is selected to meet the requirements of the job it must perform. For individual motor characteristics see DIRECT-CURRENT MOTOR; INDUCTION MOTOR; REPULSION MOTOR; SYNCHRONOUS MOTOR; UNIVERSAL MOTOR. For comparison of all ac motors, see ALTERNATING-CURRENT MOTOR.

Principles of motor operation. When a conductor located in a magnetic field carries current, a mechanical force is exerted upon it. This force has the value

$$F = 0.1\,Bil \quad \text{dynes}$$

where i is the current in amperes, B is the magnetic density in cgs lines per square centimeter, and l is the conductor length in centimeters. The illustration shows the relative directions of current, field and force. The force reverses with either current or



Relative directions of field flux, current, and force.

field reversal, but not when both are reversed. The torque T is the product of this force and the rotor radius.

If the conductor moves in the direction of F, an emf e is generated which opposes the current (motor action). If the conductor is moved against F, this emf will assist the current (generator action). Its value is

$$e = vBl \times 10^{-8} \quad \text{volts}$$

where v = velocity of conductor across the flux in centimeters per second.

The product ei represents the power converted, in watts,

motor output = ei − rotative loss,
generator shaft input = ei + rotative loss

These expressions are the bases for the emf and output formulas of dynamo machinery. Many machines will operate as either a motor or a generator, but they should be designed for the particular service.

Current may be fed into the field and armature by conduction, as in dc machines and ac series motors, or into the stator by conduction and the rotor by induction, as in ac induction and repulsion motors. [A.F.P.]

Bibliography: A. E. Fitzgerald and C. Kingsley, *Electric Machinery*, 1952; NEMA Motor and Generator Standards MG1–1955.

Motor systems

Those portions of the nervous system which regulate and control the contractile activity of muscle and the secretory activity of glands. Glands and muscles are the two types of organs by which an organism reacts to its environment; together they constitute the machinery of behavior. Cardiac muscle and to a lesser extent, some smooth muscle a

glandular structures can function independently of the nervous system but in a poorly coordinated fashion. Skeletal muscle, however, is entirely subservient to neural control. Destruction of the nerves supplying skeletal muscles leads to paralysis, an inability to move.

This article discusses skeletal muscle innervation; the motor unit; neuromuscular transmission; muscle contraction; the motor cortex; and cerebellar regulation of movement. *See* GLAND; MUSCLE; NERVOUS SYSTEM.

Skeletal muscle innervation. A skeletal or striated muscle consists of a bundle of individual elements called muscle fibers. The fibers are held together in the bundle by connective tissue and their ends are attached to movable bones so that muscle shortening during contraction results in movement. The nerve supply to skeletal muscles of the limbs and trunk is derived from large nerve cells called motoneurons (motor neurons) which are situated in the ventral horn of the gray matter of the spinal cord (Fig. 1) Innervation of muscles of the face and head originate in the motor nuclei of the cranial nerves. The axons of the motoneurons traverse the ventral spinal roots (or the appropriate cranial nerve roots) and reach the muscles via peripheral nerve trunks. Each muscle fiber in the muscle receives a nerve twig which terminates on the fiber in a specialized structure known as the neuromuscular end-plate (Fig. 2). At the end-plate the terminal twig flattens and forms an intricate arborization which is closely applied to, but is not continuous with, the muscle fiber membrane. *See* CRANIAL NERVE.

Motor unit. The number of muscle fibers in a muscle far exceeds the number of motoneuron axons in the nerve supplying it. Near the muscle each motor axon divides repeatedly, furnishing twigs for a group of muscle fibers. Activation of a motoneuron thus causes contraction of a number of muscle fibers; a single motoneuron together with the muscle fibers which it supplies is termed a motor unit. The innervation ratio, that is, the ratio between number of motor nerve fibers supplying the muscle and the number of muscle fibers in the muscle, gives a measure of the size of the motor unit. In the large, slowly contracting muscles principally concerned with postural adjustments, for example, leg muscles, the innervation ratio is estimated at from 1:120 to 1:150; that is, each motoneuron controls 120–150 muscle fibers. The motor units of the quickly contracting extrinsic eye muscles are much smaller (about 1:6). In general, muscles involved in delicate rapid movement appear to have smaller innervation ratios than do larger muscles concerned with gross movements.

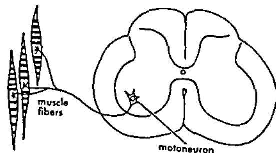


Fig. 1. Diagrammatic representation of muscle innervation. The muscle consists of many muscle fibers, only a few of which are shown here. The muscle fibers are innervated by axons derived from motoneurons, each motoneuron supplying a number of muscle fibers. A single motoneuron along with the muscle fibers which it supplies is called a motor unit, only one of which is shown in the diagram.

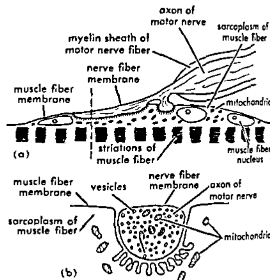


Fig. 2. Diagram of neuromuscular end-plate. (a) Structure as seen by light microscopy. Section is cut parallel to long axis of muscle fiber. (b) Structure as seen by electron micrography at much higher magnification than (a). Section is at right angles to the long axis of muscle fiber, as through the dotted line in (a). (*After R. Couteaux, Exptl. Cell Research, Suppl. 5 294, 1958, and J. D. Robertson, J. Biophys. Biochem. Cytol. 2:369, 1956*)

SKELETAL MUSCLE CONTRACTION

Neuromuscular transmission. The immediate sequence of events which leads to contraction of skeletal muscle begins with the initiation of an action potential in the motoneurons supplying the muscle. The nerve impulses propagate over the motor nerve fibers to their terminals in the end-plate. There a transmitter agent is released which in turn initiates an action potential in the end-plate membrane; this progagates over the membrane of the muscle fiber at a speed of about 5 m/sec. The muscle action potential triggers the contractile process. *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY.

The neuromuscular transmitter agent is acetylcholine which is synthesized by the nerve terminals and presumably forms the content of tiny vesicles

seen in electron micrographs of nerve terminals. The vesicles are thought to rupture at the surface to release acetylcholine. Even in the resting state small quanta of acetylcholine are released spontaneously. The action potential accelerates the rupture of vesicles and liberates enough acetylcholine to generate an action potential in the end-plate membrane.

Acetylcholine has the property of depolarizing the muscle membrane and thus initiating an action potential; it is particularly effective in depolarizing the specialized muscle membrane of the end-plate, the latter being about 1000 times more sensitive to acetylcholine than nonjunctional regions of the muscle membrane. It is estimated that a nerve impulse effective in setting up a muscle action potential does so by liberating only about 10^{-17} M of acetylcholine from the nerve terminals. The end-plate membrane contains a high concentration of an enzyme, acetylcholine esterase, which is capable of splitting, within a few milliseconds the liberated acetylcholine into pharmacologically inert degradation products, acetate and choline. The presence of acetylcholine esterase in the end-plate membrane thus rigidly limits the duration of action of the transmitter agent and prevents prolonged depolarization and repetitive firing of the muscle end-plate. *See* ACETYLCHOLINE; ENZYME.

Muscle contraction. The contractile elements of muscle consist of two proteins, actin (molecular weight about 70,000) and myosin (molecular weight about 800,000) present in a ratio of about 1:3. The complex, called actomyosin, forms long molecular chains. In the presence of adenosinetriphosphate (ATP), actomyosin filaments extracted from muscle contract and can lift weights. Each muscle fiber contains up to 10,000,000 such myofilaments of actomyosin. Older theories of muscle contraction postulated a coiling or folding of the filaments; more recent evidence based on electron microscopic studies of muscle suggests that the actin and myosin elements lie parallel to one another and that during contraction the interdigitated elements slide past each other, thus shortening the muscle (Fig 3). *See* ADENOSINETRIPHOSPHATE (ATP); MUSCLE (BIOPHYSICS).

Excitation-contraction coupling. The events intervening between the conduction of the action potential over the muscle membrane and the onset of contraction are not entirely clear. It may be that the muscle action potential causes the liberation of an agent which activates the contractile elements. A difficulty for this hypothesis is that the interval between action potential and onset of contraction is too brief (about 1 msec) to allow for diffusion of an activator from the surface membrane to the innermost myofilaments. However, electron micrographs indicate the existence of a transverse membrane, the Z line or membrane, which is continuous with the myofibrils and attached to the membrane. An activator substance may well be released from the transverse Z membrane and instigate contraction in the time interval between action potential and contraction. Significantly, when punctate electrical stimuli are applied through minute (2 μ) electrodes to single muscle fibers, localized contractions are readily produced if the electrode tip is on or near a Z line, whereas much more intense currents fail to produce contraction when the electrode is on the surface membrane midway between two Z lines.

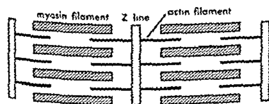


Fig 3. Diagram of arrangement of actin and myosin filaments in the myofibril as revealed by electron microscopy. During muscle contraction the actin filaments slide past the myosin filaments, drawing the Z lines closer together.

Summation of contraction. The muscle action potential, like the action potential of nerve, is an all or none phenomenon; that is, its amplitude cannot be increased by increasing the strength of the stimulus eliciting it. The contractile mechanism, however, is graded in behavior and the tension (called a twitch contraction), produced when a single muscle action potential is elicited, is considerably less than maximal. Because of this and because the contractile process greatly outlasts the duration of the muscle action potential and the concurrent refractoriness of the membrane, the tension elicited by two action potentials briefly separated in time is greater than that produced by a twitch. The twitch contraction induced by the second action potential sums with that persisting from the first action potential. If several muscle action potentials are initiated in rapid succession, the successive contractile responses sum, each subsequent contraction adding a diminishing increment of tension until further action potentials, although maintaining the tension, cannot increase it further. Such a contractile response is called a tetanus. In some muscles the tetanic tension is nearly four times as great as the twitch tension. The rate of discharge necessary to produce maximal tetanic tension varies in different muscles from 30 discharges/sec for slow soleus muscle to 350 discharges/sec for the rapidly contracting internal rectus muscle of the eye.

Gradation of muscle contraction. In order to participate smoothly in a variety of motor acts, muscular contraction must be graded in accordance with the role of the muscle in each act. The motor unit is the smallest functional unit of muscle, and therefore the weakest possible natural movement is the twitch of a single motor unit. It is estimated that such a contraction may generate a tension of about 10 g in the medial gastrocnemius muscle of

the cat. In voluntary contractions progressively greater tensions are produced by the following mechanisms: (1) more motor units discharge (recruitment), (2) active motor units discharge at more rapid rates but not rapidly enough for contractile summation, and (3) further increase in frequency of motor unit discharge leads to summation of individual motor unit twitches and consequently to tetanic tensions. In some muscles during voluntary contraction the rate of motor unit discharge increases from 5 to 50 action potentials/sec as the contraction increases from light to maximal effort. At the lower rates summation does not occur; each motor unit produces a series of unfused twitches. However, because the motor units comprising one muscle discharge asynchronously, the total tension in the muscle varies with the average frequency of discharge; thus even through the range of subtetanic frequencies the contraction is smoothly graded. At higher frequencies of discharge summation of individual motor unit twitches produces still further increase in tension.

Regulation of motoneuron discharge. Motoneurons are discharged by nerve impulses reaching them through a variety of neural pathways. Some of these impulses originate in sense organs in the skin and subcutaneous structures and reach the spinal motoneurons via the dorsal spinal roots. In many instances these pathways are fairly direct and can be readily traced as a chain of neurons from sense organs to muscle. Such simple segmental paths constitute reflex arcs. For a description of reflex arcs, *see* REFLEX, UNCONDITIONED.

Other chains of neurons impinging on motoneurons are so complex that complete arcs cannot be traced, the number of neurons and the possible number of connections with peripheral inputs being so enormous as to defy unequivocal analysis. Some systems of this type may, however, be traced backward over several neurons up the neuraxis as far as the brain stem or cerebral cortex where the connections with fibers bearing incoming impulses are so multitudinous that the trail is lost. There is every reason to suppose that the so-called cortical and subcortical motor paths merely constitute part of complex interrelated long chain arcs; the term "cortically originating motor pathways" is really a cloak for ignorance.

MOTOR CORTEX

The cerebral *cortex* is a mantle of nerve cells covering the forebrain. The axons of some of these *cells pass into* the underlying white matter and descend through *the* brain stem to reach directly or indirectly the motoneurons *of the* brain stem or spinal cord. Areas of cortex in which such *corticifugal cells* are numerous are designated motor cortical areas. *Operationally* the motor areas are defined as cortical areas in which electrical stimula*tion elicits movement* and excision produces either paralysis or *paresis (weakness)*.

Areas defined. *Punctiform* exploration of the exposed cerebral cortex with electrical stimulation reveals several prominent motor areas. One of these is in the anterior wall of the central fissure of Rolando, and in the monkey it extends over the precentral gyrus; in man the precentral motor area is for the most part buried in the central fissure. Weak repetitive stimulation within this region elicits discrete movement, for example, a slow flexion of the thumb or a retraction of the lip. The evoked movement is always on the side of the body contralateral (opposite) to the stimulated hemisphere; the corticifugal pathways (that is, those pathways leading away from the cortex) cross the neuraxis before reaching the spinal cord. Moreover, the muscles activated vary with stimulus site within the precentral motor area. Beginning at the longitudinal fissure and proceeding laterally and downward along the precentral gyrus, one finds an orderly array of motor points for leg, trunk, arm, and face musculature. The topographical array may be represented by displaying the body superimposed upside down along the precentral gyrus (Fig. 4). In addition to this vertical topographical organization a frontoposterior organization is evident so that the musculature of the apical portions of the body (fingers, toes, lower lip, and tip of tongue) are posteriorly located, mostly in the depths of the central fissure, whereas the musculature of trunk and back is represented more rostrally on the free surface of the precentral gyrus. The organization is represented in *Figure* 3 by a figurine standing upside down on *the cortex* with toes, fingers, tongue, and lower lip wedged in the gutter of the central fissure. Another striking fea-

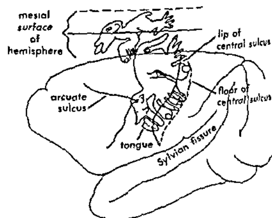


Fig. 4. Figurines showing motor representation in precentral motor area (lower figurine) and supplementary motor area (upper figurine) of the monkey. *Parts* of the body are superimposed on cortical *areas* which regulate their movement. The *central sulcus* is opened to show representation in the buried parts of the fissure. Similarly in the upper part of the *figure* the mesial surface of the hemisphere is *folded* back to show the organization of motor points ordinarily hidden in the longitudinal fissure separating the two hemispheres. (*After* C. N. Woolsey *et al*, *Research Publs. Assoc. Research Nervous Mental Disease*, 30 *238–264, 1952*)

ture of motor cortex organization is the disproportion of cortical area devoted to different muscles. The muscles of fingers, toes, lips, and tongue, that is, portions of the musculature involved in delicate, precise movements, have relatively large cortical areas devoted to their control. In the monkey the thumb and great toe which are involved in grooming, precise manipulation, and grasping have especially extensive cortical representation. On the other hand, relatively small cortical areas are devoted to trunk musculature which is principally concerned with grosser movements. The disproportion of cortical representation of different muscles gives rise to the distortion of the figurine in Fig. 3.

More precise mapping of the precentral motor area indicates that even individual muscles have more or less punctate representation. Muscle representation is not entirely discrete, for there is a considerable overlap, but for each muscle cortical foci are found where electrical stimulation most readily elicits contraction. In terms of cellular organization of the cortex the picture is one of an overlapping mosaic of cell groups, each group devoted to driving the motoneurons supplying individual muscles. Electrical stimuli, particularly strong ones which spread through a considerable volume of cortex, discharge cortical cells indiscriminately so that the pattern is obscured unless great care is taken. In the natural state, presumably ascending impulses arriving at the cortex play discriminately upon the motor cortex, activating the motor points required to bring into play the various muscles involved in a particular movement, much as the pianist makes different chords by appropriate selection of certain keys on the keyboard.

Although the precentral motor area is the best known motor cortical area, two other motor areas for the body musculature have been described. One of these, called the second motor area, is said to be buried in the Sylvian fissure; its organization is face to face with the precentral motor representation. Because of its limited accessibility this area has not been thoroughly mapped. Another complete motor area, the supplementary motor area, lies for the most part on the mesial surface of the hemisphere. The extent of the area and its topographical organization (which is not as clearly defined as that of the precentral motor area) are shown in Figure 4. The threshold to electrical stimulation is higher and more readily affected by anesthetics for supplementary motor area than for precentral motor area. Also, the movements initiated by supplemental motor area stimulation tend to persist may seconds after stimulation has ceased, in contrast to the phasic movements elicited by precentral motor stimulation. Finally, supplementary motor area responses often involve the musculature of both sides of the body.

It thus appears that the musculature of the entire body is controlled by at least two, and possibly three, discrete sets of cortical neurons.

Eye and neck movements. Stimulation of the cortex in the region anterior to the arcuate fissure elicits movements of neck and eyes. The typical response consists of conjugate movement of the eyes and turning of the head to the side opposite the stimulated hemisphere. Often the movement is associated with autonomic motor responses of the eyes, for example, lacrimation and pupillary dilation and constriction. For this reason the area enclosed between the two limbs of the arcuate fissure is often referred to as the frontal eye field. Whether it constitutes a separate motor area or merely an extension of the precentral motor area is not entirely clear. The significance of the coordinated head and eye movements elicited by stimulating the eye fields is made clearer when it is recalled that all the major ascending pathways leading into the cortex cross the neuraxis prior to reaching the cortex. Thus stimuli applied to the right side of the body generate nerve impulses which reach the left hemisphere. These impulses activating the motor eye fields of the left hemisphere cause movement of the head and eyes toward the right, that is, toward the stimulated side of the body. Such orientational movements may be part of the motor aspect of attention.

Response from other cortical sites. The cortical areas described above constitute the traditional motor areas as defined by electrical stimulation experiments. Cortical stimulation of other areas, especially in lightly anesthetized or unanesthetized animals may elicit movements. In experiments on unanesthetized monkeys with 610 stimulating electrodes permanently implanted on the lateral surface of the cortex, it was found that stimulation at nearly any point gave rise to movement, although the type of movement was highly variable from point to point. Evaluation of such experiments is complicated by the fact that the motor areas receive axons from intracranial association neurons situated in other cortical areas, and consequently elicitation of movement by cortical stimulation does not necessarily betoken corticifugal projection from the stimulated point.

Effects of ablations. Removal of motor cortex within and just anterior to the central fissure causes a loss of voluntary movement in the contralateral muscles represented in the ablated cortex. Lesions confined to the leg representation cause paralysis of the leg, leaving the arm and face musculature unaffected; thus the topographical organization of motor cortex is revealed both by stimulation and ablation experiments. The paralyzed extremity is flaccid; that is, it hangs limply and displays no resistance to passive flexion or extension. Paralysis of voluntary movement is not permanent, however; the duration and severity of deficit varies with species, being more severe in man, ape, and monkey than in lower mammals such as dog and cat. This species difference reflects the increased dominance of the cerebral cortex in phylogenetically advanced forms. Following ablation of the arm [illegible] cortex in monkeys and chimpanzees, mov[illegible] at the shoulder begins to reappear withi[illegible] days [illegible] ater, movements at elbow and [illegible]

pear, and finally (8–12 weeks in chimpanzees) crude movements of fingers are restored. However, finger movements never regain the preoperative delicacy; thumb-finger approximation such as occurs in grooming remains permanently defective and awkward. This is a reflection of the extensive cortical control of the finger muscles which have a large cortical representation.

When the ablation is extended farther rostrally to include the more anterior parts of the precentral gyrus, paralysis is again evident in voluntary movement, but the paralyzed limb becomes stiff rather than flaccid and the experimenter's attempt to move the limb at the joints meets with resistance. The resistance is due to active muscle contraction which opposes movement of the joint. Such a paralysis is spastic. Despite the active involuntary contraction of the muscle, voluntary movements are weak or lacking

Following isolated unilateral ablation of the frontal eye fields in monkeys, the animal's head and eyes turn toward the side of the lesion; there is transient paralysis of conjugate deviation of the eyes toward the opposite side. In walking, the animal tends to turn toward the side of the lesion; that is, it tends to follow the eyes and head, so that locomotion is circular. The circling gait may persist after paralysis of the eye muscles has disappeared.

Pyramidal tract. One of the best-known corticifugal pathways is the pyramidal tract. The anatomy and function are discussed in this section.

Anatomy. The pyramidal tract is the only tract of nerve fibers which originates from cortical cells and runs without interruption through the brainstem to reach the spinal cord; for this reason it is also known as the corticospinal tract. Just before reaching the spinal cord the tract runs along the inferior surface of the medulla where it forms a grossly recognizable band of fibers known as the pyramid, hence the name pyramidal tract. The human medullary pyramid contains about 1,000,000 fibers, mostly of small diameter.

The axons of the pyramidal tract originate entirely from cortical cell bodies; after complete decortication, the entire pyramid degenerates. About 40% of the total number of fibers originate from cell bodies in the precentral motor area. Some of these fibers arise from the giant pyramidal cells of Betz, large conical cells which constitute the major histological distinction of precentral motor cortex In man, the motor area of each hemisphere [illegible]

[illegible] fibers originating from precentral motor cortex stem from smaller, less distinctive cells, located mostly in the deeper layers of the cortex. The supplementary motor area contributes no fibers to the pyramidal tract but sends into the precentral motor area a heavy projection of fibers which form connections with the pyramidal tract cells of origin located there.

Another component of small pyramidal fibers derives from the postcentral gyrus just posterior to the central fissure. The postcentral contribution is estimated at about 20% of the total. Together precentral and postcentral areas account for only about 60% of the fiber content of the pyramid. The origin of the remaining 40% is not clear. Some fibers apparently arise from the cortex anterior to the precentral motor area, but the size of the contribution is not known.

The axons of the pyramidal tract leave their cortical cells of origin to enter the subcortical white matter and thence pass through the internal capsule, the cerebral peduncles, and the pons (Fig. 5). Some fibers terminate in the brain stem on motoneurons of cranial nerve nuclei; by definition these fibers are neither corticospinal nor pyramidal, although they presumably have the same function as their neighbors, which are destined for spinal rather than cranial motoneurons. At the pontomedullary junction the tract emerges on the ventral surface of the medulla as the pyra- [illegible]

[illegible] gled with fibers of other [illegible] nuclei.

[illegible] only through this short segment of their course [illegible] the corticospinal fibers are found in "pure culture" isolated from other systems.

At the caudal border of the medulla the pyramids plunge dorsally into the substance of the cord and most of the fibers cross to the opposite side of the

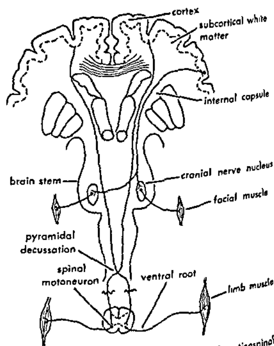


Fig. 5. Diagram of origin and course of corticospinal (pyramidal) tract. Fibers to cranial motor nuclei (sometimes called corticobulbar fibers) leave tract before it enters the pyramid just above the decussation.

neuraxis to occupy a position in the lateral column of the spinal cord. Below the crossing point, or decussation the corticospinal fibers are again mixed with fibers of other tracts. Some pyramidal fibers do not decussate but continue either in the ventral white funiculi or in the lateral column of the same side. The size of the uncrossed pyramidal component varies in different species. The spinal portion of the corticospinal tract extends throughout the full extent of the spinal cord but progressively diminishes in size as it passes from cervical to lumbar segments because many fibers terminate in the gray matter of all segments. Terminations are particularly numerous in the cervical and lumbar enlargements which contain the motoneurons supplying the musculature of upper and lower limbs. Some corticospinal fibers make direct connections with motoneurons; others terminate in the dorsolateral and intermediate gray matter on cells whose axons connect directly or indirectly with motoneurons.

Functions. The role of the pyramidal tract in regulating motor function in monkeys and chimpanzees has been studied by sectioning the medullary pyramid on one side and observing the resulting changes in motor performance. The most prominent defect following unilateral pyramidotomy is contralateral paresis involving the musculature from the neck down. The affliction is more severe in chimpanzees than in monkeys; in the former even stereotyped movements of progression are impaired but not abolished. In neither animal is paralysis so grave as to render the affected parts useless, but there is severe impairment of voluntary movement and loss of such fine movements as opposition of thumb and index finger in grooming and individual movements of the fingers in manual exploration. This type of deficit has been observed to persist up to 4 years after operation, and thus may be considered permanent.

Associated with the paresis, especially in monkeys, is flaccidity reminiscent of that seen following precentral cortical ablations; the extremities are limp and offer no resistance to movement at the joints. It is clear that the spasticity resulting from rostral, precentral lesions (and in man from hemorrhage into the internal capsule) is due to interruption of some corticifugal paths other than the pyramidal tract.

Extrapyramidal systems. Isolated destruction of the corticospinal tract does not cause complete paralysis resulting in a total loss of voluntary movement Furthermore, electrical stimulation of the motor cortical areas after pyramidotomy still elicits muscular contraction. It follows that the cortex must give rise to motor pathways other than the pyramidal tract. These other pathways are often referred to collectively as "cortically originating extrapyramidal pathways." This is an anatomical term of convenience but from the physiological point of view an undesirable one because there is no reason to believe that the extrapyramidal systems are functionally identical. None of the extrapyramidal systems pursues an uninterrupted course from cortex to spinal cord; rather, impulses are relayed through various brain-stem nuclei and reach the spinal levels only after multiple synaptic relays. Consequently both the anatomical and the functional study of extrapyramidal systems is complicated. There is reason to suppose that extrapyramidal projections arise from nearly all portions of the cortex, although the rostral part of the precentral motor area appears to give rise to particularly large contributions.

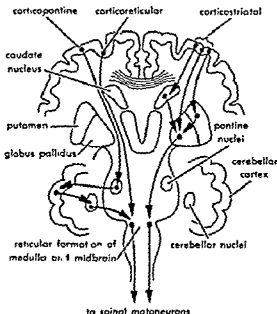


Fig 6. Diagram of extrapyramidal projection systems.

Some of the better known extrapyramidal systems (Fig. 6) may be briefly listed.

1. Corticostriatal and corticopallidal systems originate from rostral precentral cortex and project to the caudate nucleus and the putamen, which are portions of the basal ganglia. From the basal ganglia impulses are relayed to the midbrain where they presumably relay to spinal levels.

2. Corticoreticular systems originate from the cortex around the central fissure, especially from the motor area. The axons terminate in the region of the pons and medulla on the diffusely organized neurons constituting the pontine and medullary reticulum. The projection is bilateral and poorly, if at all, organized somatotopically. Impulses are presumably relayed to the spinal cord via the reticulospinal tracts which traverse the ventral and lateral white columns of the cord. *See* RETICULAR FORMATION (BRAIN).

3. Corticopontine systems arise from the cortex of each of the four lobes of the brain although the contribution from the rostral precentral gyrus is most prominent. A significant contribution also comes from the supplementary motor cortex on t mesial surface of the hemispheres. The fiber

minate in the pontine nuclei which in turn project to the cerebellum. From the cerebellum, projection systems feed impulses into the medullary reticulum, which in turn projects through the reticulospinal tracts to the spinal cord as described above.

CEREBELLAR REGULATION OF MOVEMENT

The cerebellum has no direct pathways connecting it with the spinal motoneurons, but it exerts control over movement via paths which originate in the cerebellar nuclei and play upon certain brainstem nuclei (red nucleus, vestibular nuclei, and reticular formation) and upon the precentral motor cortex. Cerebellar injury causes no paralysis in the real sense but it does result in marked deficiencies in the delicacy and accuracy of voluntary movement and produces disturbances in postural coordination.

Afferent inputs. The cerebellum receives a bewildering variety of afferent inputs, the most important of which are (1) the spinocerebellar tract which originates in the spinal cord and receives impulses from the primary afferent fibers supplying the muscle spindles, (2) a vestibulocerebellar projection from the vestibular nuclei, and (3) corticopontocerebellar tracts which connect the cerebellum with all four lobes of the cerebral cortex.

Function. The function of the cerebellum has been likened to that of a servocontrol mechanism which detects and corrects errors during the course of a movement (*see* SERVOMECHANISM). Following injury to the cerebellum, movements are ataxic; that is, they [illegible] and di[illegible] directi[illegible] in reaching for an object the hand often overshoots or undershoots the target. Voluntary movement is characterized by tremor which increases as the movement progresses. All these disturbances appear to be due to defects, not in executing movement, but in detecting errors and instigating correction of movement. Thus, if in reaching for an object the hand strays from the correct path, correction is begun too late and proceeds too far, so that the hand overshoots to the other side of the direct path. A succession of such faulty corrections gives rise to the oscillating tremor typical of cerebellar injury. The cerebellum, which receives sensory information concerning the position of the limb in space from the muscle spindles and which has both incoming and outgoing connections with cortical and subcortical centers responsible for movement, is essential for coordination of movement. The exact neuronal mechanisms by which this delicate control is accomplished are not known.

[illegible] Te[illegible] pr[illegible] ga[illegible] the victim, although showing little incoordination of movement while lying in bed, is unable to maintain balance and coordination of limb musculature in walking.

Another disturbance of muscular coordination following destruction of certain parts of the cerebellum is a reduction of muscle tone (*see* POSTURE, REGULATION OF). The limbs show less than normal resistance to passive movement at the joints. The mechanism of cerebellar hypotonia is not clear. *See* PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL. [H.D.P.; T.C.R.]

Motorboating

A form of oscillation that occurs at a very low audio frequency in a system, circuit, or component. It is caused by excessive amount of audio feedback at low frequencies. This oscillation is a succession of pulses; when these occur in a circuit that is feeding a loudspeaker, the pulses result in putt-putt sounds resembling those made by a motorboat.

Feedback through a common power supply is one cause of motorboating; this can be suppressed by using resistance-capacitance decoupling filters between the power supply and the plate circuit of each tube. Another solution is to cause the amplification to fall off sharply immediately below the useful range of frequencies, so that the motorboating frequencies are inherently suppressed. *See* AMPLIFIER. [J.M.R.]

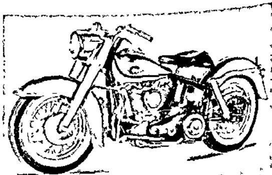
Motorcycle

A wheeled, trackless, self-propelled vehicle for land transportation. A 2-wheeled model as illustrated carries a driver, or driver and passenger. A 3-wheeled model having two wheels at the rear is used for package delivery and for automobile pickup and delivery.

The driver starts the engine by pulling a cord or stepping on a ratchet lever to rotate the crankshaft. He then controls engine speed and the clutch, shifts gears, and applies brakes by twisting grips on the handlebars and by foot pedals.

A motorcycle is rated by the cubic-inch displacement of its engine and further identified by type. There are auxiliary-engined bicycles, scooters, light-weights, and heavy-weights.

The engine is slung in a frame and has one or two cylinders. The two cylinders are in-line, in V formation, or horizontally opposed. Engines oper-



Heavy-weight model motorcycle with V-type 2-cylinder engine, chain drive to rear wheel, helical spring front suspension, and swing-arm 3-point rear suspension. (Harley-Davidson Motor Co.)

ate on the 2-stroke or 4-stroke cycle and are air-cooled.

Normally, a multiple-disk clutch is interposed between engine and transmission. The transmission has three or four speeds forward, and a reverse gear if the vehicle has three wheels. Power is transmitted from the transmission by shaft, gears, or by roller chain to a sprocket on the rear wheel. In some scooters, power is transmitted by a belt running over two pulleys. The pitch diameter of the pulleys is changed by centrifugal force to alter the ratio between engine speed and wheel speed, thus eliminating both clutch and transmission and providing an infinitely variable drive.

Brakes of the internally expanding type are fitted to the front and rear wheels.

The driver's saddle is cushioned on coil springs. The front wheel is sprung by means of long helical springs in the front fork, hydraulically damped. The rear wheel may or may not be sprung. On heavy-weight models as illustrated the rear suspension consists of swing-arms sprung by means of two helical coil springs which are controlled by two hydraulic shock absorbers. Coupling this with a centrally pivoted spring on the frame gives a 3-point suspension which permits the rear wheel to move in a vertical plane independently of the frame. See AUTOMOTIVE VEHICLE. [P.H.S.]

Motor-generator set

A motor and one or more generators, with their shafts mechanically coupled, used to convert an available power source to another desired frequency or voltage. The motor of the set is selected to operate from the available power supply; the generators are designed to provide the desired output. Motor-generator sets are also employed to provide special control features for the output voltage.

The principal advantage of a motor-generator set over other conversion systems is the flexibility offered by the use of separate machines for each function. Assemblies of standard machines may often be employed with a minimum of engineering required. Since a double energy conversion is involved, electrical to mechanical and back to electrical, the efficiency is lower than in most other conversion methods. In a two-unit set the efficiency is the product of the efficiencies of the motor and of the generator.

Motor-generator sets are used for a variety of purposes, such as providing a precisely regulated dc current for a welding application, a high-frequency ac power for an induction-heating application, or a continuously and rapidly adjustable dc voltage to the armature of a dc motor employed in a position control system. See GENERATOR, ELECTRIC; MOTOR, ELECTRIC. [A.R.E.]

Mountain

A feature of the earth's surface that rises high above its base, has generally steep slopes and a relatively small summit area. Commonly the features designated as mountains have local heights measurable in thousands of feet, lesser features of the same type being called hills, but there are many exceptions.

Mountains rarely occur as isolated individuals. Instead they are usually found in roughly circular groups or massifs, such as the Olympic Mountains of northwestern Washington and the Harz Mountains of northern Germany, or in elongated ranges, like the Sierra Nevada of California, the Bighorn Range of Wyoming, or the Sierra de Guadarrama of central Spain. An array of linked ranges and groups, such as the Rocky Mountains, the Alps, or the Himalayas, is a mountain system. North America, South America, and Eurasia possess extensive cordilleran belts, within which the bulk of their higher mountains occur. See CORDILLERAN BELT; MASSIF; MOUNTAIN SYSTEMS.

As a rule, mountains represent portions of the earth's crust that have been raised above their surroundings by upwarping, folding, or buckling, and have been deeply carved by streams or glaciers into their present surface form. Some individual peaks and massifs have been constructed upon the surface by outpourings of lava or eruptions of volcanic ash. See VOLCANO. [E.H.HA.]

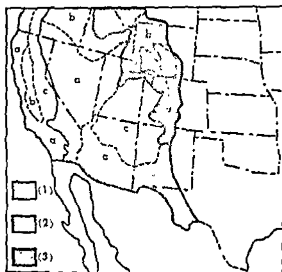
Mountain systems

Major divisions of a cordillera and its several ranges or other topographic elements tied together by common geologic features that set a system apart from adjacent ones. The systems of the North American cordillera number about 20 depending on various classifications. Exemplification may be made in the western United States where the cordillera is most complex. There the systems from California eastward are (1) the Pacific Mountain systems, composed of the Coast Ranges, the Great Valley, and the Sierra Nevada; (2) the Intermontane Plateau system comprising the Basin and Range system (Great Basin) with the Columbia and Colorado Plateaus; and (3) the Rocky Mountain systems, including the Southern, or Colorado, Rocky Mountains, the Wyoming Basin, and the Wyoming, or Middle, Rockies. See TECTONIC PATTERNS.

Pacific mountain systems. This western division presents one basin and two mountain divisions.

Coast Ranges. These are composed chiefly of folded and faulted Cretaceous and Cenozoic sedimentary rock. They are very young mountains, and the incessant earthquake activity indicates that vigorous growth is continuing at the present time.

Sierra Nevada. A great tilted fault block, this is uplifted on the east and is composed dominantly of granite and older metamorphosed strata. The tilting is of recent geologic age, like the deformation in the Coast Ranges, and is probably continuing today. The granite, however, attests to an earlier episode of mountain building (early Cretaceous) when great volumes of molten rock rose from depths and invaded the upper layers of the earth's crust, there to cool and crystallize. The vast ir



Sketch map of mountain systems of southwestern United States. (1) Pacific Mountain system: (a) Coastal Ranges, (b) Central Valley (basin), (c) Sierra Nevada; (2) Intermontane Plateau systems: (a) Basin and Range (Great Basin), (b) Columbia Plateaus, (c) Colorado Plateaus; (3) Rocky Mountain systems: (a) Southern Rocky Mountains, (b) Middle Rocky Mountains, (c) Wyoming Basin.

sions (batholiths) are approximately aligned in a belt which extends from the Aleutian and Alaska ranges through the Coast Ranges and island archipelago of southeastern Alaska and British Columbia to the northern Cascades and the Idaho batholith to the Sierra Nevada. Southward the belt forms Baja California and probably the Sierra Madre del Sur of southern Mexico.

Intermontane Plateau systems. Three basin and plateau systems stand between the Pacific and Rocky Mountain systems.

Basin and Range systems. This is an upland of geologically young fault blocks, some uplifted and uptilted to form the northward trending block-mountain ranges, and some down-dropped to form the intermontane valleys or desert bolsons. The internal architecture of the ranges attests an older stage of folding and overthrusting of the strata. Large areas of the Great Basin are masked with young volcanic rocks, some older than the block faulting, and some younger.

Columbia Plateaus system. Mainly built up of nearly horizontal sheets of lava, the generally flat to rolling upland surface is marked here and there with notably varied features. A few streams in the dominantly dry region have cut valleys and canyon sections; and some mountains surmount the nearly continuous lava uplands in patterns from simple to complex.

Colorado Plateaus system. A few great steplike uplands are a relatively undeformed section of the earth's crust. The flat-lying sedimentary rocks are occasionally deeply dissected to form canyons as a conspicuous characteristic.

Rocky Mountain systems. Two mountain systems and an intervening basin extend from Wyoming only as far as northern New Mexico.

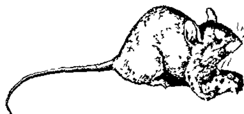
Southern Rocky Mountains. These are a complexly compressed belt of mountains with occasional high intermontane basins (Colorado "parks").

Middle Rocky Mountains. Eroded complex rock structures present anticlinal remnants in many of the present ranges. These also include several intermontane basins.

Wyoming Basin. Elevated remnants of plains stand, in varying degrees of current erosion, between the Middle and Southern Rocky Mountains The basin plains are surmounted by scattered low mountains. [A.J.E.]

Mouse

In its broadest sense the name mouse is used to designate all small rodents. Typically, mice are short-haired and have prominent ears and long tails. Of the wide variety of Cricetidae, the New World mice and rats, the North American genus *Peromyscus* is the most successful. These are known as the white-footed mice. There are several species and numerous subspecies in the United States.



The house mouse, *Mus musculus*; length to 7 in. (From E. L. Palmer, Fieldbook of Natural History, McGraw-Hill, 1949)

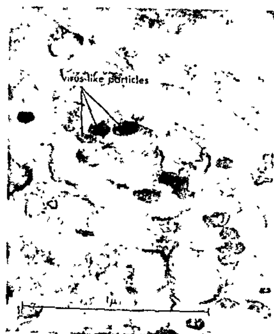
The mice of the subfamily Microtinae, family Cricetidae, are short-tailed, short-eared, stoutly built animals. Collectively, they are often called voles, but are also known as meadow mice, *Microtus*, and lemmings. The genus *Microtus* includes the most prolific of all vertebrates, sometimes producing amazingly large populations. These mice are the most destructive of all small rodents.

The Old World rats and mice, family Muridae, include both the common house mouse and the Norway rat, two of man's most destructive rodent enemies. Both were accidentally introduced into the United States.

Although mice are generally terrestrial animals, living in surface or underground nests, a number are arboreal. Most are herbivorous, but many species also utilize animal food to supplement their diet. See LEMMING; RAT; RODENTIA. [J.D.S.]

Mouse viral leukemias

A filterable agent (virus) has been obtained from a strain of mouse that develops leukemia spontaneously. The virus will cause leukemia in a non-



Mouse spleen, showing viruslike particles as seen through the electron microscope.

leukemic strain of mouse if the inoculations are made before the mice are 16 hours old. The fluid from mouse-embryo tissue cultures inoculated with the same material will produce a wide variety of malignant tumors in mice.

Chloroleukemia, multiple tumors of the bone marrow and soft tissues near bone, has also been reported after inoculation of filtrates of Ehrlich carcinoma. All the above types depend on inoculation of very young mice and have a latent period of 9 months. Another leukemia can be produced by inoculation of filtrates of spleens from leukemic animals into adult mice. In this disease the spleens become palpable in 2 weeks and the animals die in 3 months. It is possible to vaccinate against this virus with a formalin-killed suspension of the virus. *See* Animal virus; Leukemia; Tumor viruses.

[A.E.M.]

Mouth

In man, the oral or buccal cavity and its related structures. The margins of the lips mark the junction between the outer skin and the inner mucous lining of the gastrointestinal tract. The mucosa of the mouth forms its lining, forms the gums surrounding the teeth, and covers the surface of the tongue. *See* Gastrointestinal tract.

The roof of the mouth consists of the hard palate, and behind this, the soft palate, which merges into the oropharynx. The lateral walls consist of the distensible cheeks, and the floor is formed principally by the tongue and the soft tissues that lie between the two sides of the lower jaw, or mandible.

The teeth project from sockets, or alveoli, in the upper maxilla and the lower mandible.

Three pairs of salivary glands empty their contents into the mouth. These are the parotid, submaxillary, and sublingual glands.

The posterior limit of the oral cavity is marked by the fauces, an aperture which leads to the pharynx. On either side of the fauces are two muscular arches, covered with mucosa. These are the glossopalatine and pharyngopalatine arches between which lie masses of lymphoid tissue, the tonsils. Suspended from the posterior portion of the soft palate is the soft retractable uvula.

[E.C.ST.]

Mouth disorders

Disorders of the lips, teeth, oral cavity, and related structures.

Dental disorders are most frequent and include toothache, dental caries, abscesses, pyorrhea, and gingivitis. *See* Tooth disorders.

Salivary gland disease may give rise to ptyalism, a hypersecretion of saliva. Common causes are inflammations from drugs, poisons, and local disturbances such as irritation and reflex stimulation. Certain disorders of the nervous system may also produce excess stimulation. Diminished salivation, or aptyalism, is due to a variety of causes. Congenital defects of the salivary glands are uncommon and specific infections are infrequent, but cyst or tumor formation is often encountered, particularly in the parotid glands which are also susceptible to the everyday virus disease, mumps.

Glossitis is acute or chronic inflammation of the tongue and may be produced by a local disorder such as infection, irritation, trauma, excessive use of certain foods or beverages, and also as a result of sensitization to certain materials in food or other substances placed in the mouth.

Systemic or generalized causes of glossitis include avitaminosis (particularly of the B group), certain anemias, a few skin diseases, and infections like scarlet fever, syphilis, or tuberculosis. The tongue is said to be the mirror of disease and therefore reflects many types of disease in the body. Glossitis may also be a restricted form of stomatitis. *See* Scarlet fever; Syphilis; Tuberculosis; Vitamin.

Stomatitis is an inflammation of the mouth, especially the mucous membrane lining, produced by much the same agents that cause glossitis. Among the more commonly encountered specific infections are thrush (moniliasis), trench mouth (Vincent's angina), and canker sores (ulcerative stomatitis). *See* Candidiasis; Vincent's angina.

Other disorders of the mouth include congenital malformations of jaws, teeth, or soft tissues, inflammations from specific organisms, and neoplasia, represented by both cysts and tumors. *See* Oncology.

Except for those tumors affecting the jaws, most tumors occurring in the oral region are benign,

but premalignant or malignant conditions occur. Common benign tumors of the oral cavity are fibromas, papillomas, and hemangiomas. Malignant tumors of the oral cavity comprise about 5% of all malignant disease in humans, and nine-tenths of these are epidermoid carcinomas, with most frequent involvement of the lower lip, followed by the tongue and other structures. The remaining one-tenth includes adenocarcinomas, malignant melanomas, and other rare and less common malignancies.

Diseases of the jaws are common and include cysts, tumors, and reaction of the bone to systemic diseases such as endocrine disturbances and certain infrequent bone disorders

The peculiar structures of the mouth are susceptible to many conditions and agents because of their location, easy accessibility to irritants, and the intimate relationship of the mouth to the respiratory and gastrointestinal systems. In addition, the oral structures almost always show manifestations of certain systemic states or diseases such as pregnancy, diabetes, blood disorders, and hormonal or nutritional disturbances. *See* DIABETES; HEMATOLOGIC DISORDERS; MOUTH; PREGNANCY.

[E.C.ST.]

Moving-target indication

A method of presenting pulse-radar echoes in a manner that discriminates in favor of moving targets and suppresses stationary objects. MTI is almost a necessity when moving targets are being sought over a region from which the ground clutter echoes are very strong. The most common presentation of the output of a radar with MTI is a plan-position indicator (PPI) display. The moving targets appear as bright echoes, while ground clutter is suppressed 25 db or more. Highways and airport runways appear as lines of moving spots because of the vehicles on them.

MTI is based upon the use of the Doppler effect; that is, the carrier frequency of the echo from a target moving towards or away from the radar *shifts* by an amount proportional to the product of radial velocity and transmitted frequency. A stable *oscillator* in the receiver is synchronized with the transmitter, providing a continuous reference of the transmitter frequency and phase. In effect, it serves as a *continuation* of the carrier after the transmitter pulse is sent out. Echoes are heterodyned with *the reference* oscillator. Stationary objects supply echoes having a carrier phase shift which *is constant* from pulse to pulse, because the time for the signal to travel to a stationary object and back is always *the same*. Therefore the heterodyner output for stationary echoes does not change in phase from pulse to pulse. However, the phase shift in echoes *from* moving targets changes from pulse to pulse because of the change in the signal *propagation time*, causing the heterodyner output to exhibit a corresponding pulse-to-pulse phase difference. *See* DOPPLER RADAR.

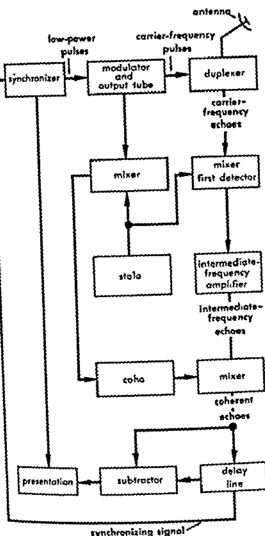
In order to utilize the lack of pulse-to-pulse phase difference to discriminate against clutter echoes, the heterodyner output is fed to a delay line which stores the signal for a period of time exactly equal to the period between pulses. Then the output of the delay line is subtracted from the freshly produced output of the heterodyner. If the two outputs are identical, as will be the case for clutter, the difference is zero. But this difference is not zero for moving targets because of their pulse-to-pulse phase shift. It is the difference between consecutive echoes that is presented as the MTI output.

The delay line usually employed utilizes acoustic propagation within fused quartz. The signal is applied to a quartz crystal which vibrates because of the piezoelectric effect. The crystal is bonded firmly to a slab of fused quartz in which the crystal vibration launches an acoustic wave. The quartz slab has a polygonal perimeter, with the sides of the polygon oriented so that the wave is reflected from one side to another, finally impinging on the side to which the output crystal is bonded. The acoustic wave causes the crystal to vibrate, creating a piezoelectric voltage which constitutes the delay-line output signal. The fused-quartz slab is shaped as a polygon of about 15 sides, such that the total internal propagation time of the acoustic wave equals the radar interpulse period. Since the delay within the quartz is difficult to vary, the pulse repetition rate of the transmitter is usually determined by and obtained from the delay line.

A variety of internal arrangements are in use for assuring that the reference oscillator is synchronized with the transmitter output. The reference oscillator may be at the carrier frequency but usually operates at an intermediate frequency, because better stability can be obtained in a low-frequency oscillator. This arrangement requires that the local oscillator used for conversion from the carrier frequency to the intermediate frequency be extremely stable, and the name stalo (stable local oscillator) is used to denote circuits that satisfy this requirement. The reference oscillator is called the coho (coherent oscillator) to denote that it remains coherent (fixed phase relation) with the transmitter output.

If the output stage is an amplifier, it can be excited from the coho via a mixer supplied from the stalo. On the other hand, if the output stage is a free oscillator, a portion of the output pulse is mixed with the stalo output and fed to the coho to synchronize it as shown in the illustration. The coho then must be resynchronized by each output pulse.

The figure of merit of an MTI system, clutter suppression factor, is defined as the ratio of rms clutter amplitude before MTI to rms clutter fluctuation after MTI. It is the amount by which the clutter is suppressed. It depends, in turn, upon several other factors, some of which are system parameters, and therefore controllable, and



Block diagram of a radar system employing moving-target indication.

some of which are functions of the terrain and weather. The uncontrollable factors are those which may cause the objects producing clutter to be nonstationary, such as the wind making foliage wave or producing whitecaps and spray on the sea. The system parameters have the following effects: (1) increasing the carrier frequency deteriorates clutter suppression by increasing the Doppler effect for moving clutter objects, (2) increasing the pulse repetition rate improves clutter suppression by reducing the distance clutter objects can move between pulses, and (3) increasing the antenna scan speed or reducing the antenna beam width deteriorates clutter suppression by changing the antenna gain in a given direction more rapidly and introducing a new clutter region for cancellation sooner; this deterioration can be substantially mitigated by using a limiting amplifier which minimizes the effect of antenna gain variation. *See* RADAR.

[R.I.B.]

Mucilage

A naturally occurring, high-molecular-weight (200,000 and up), organic plant product of unknown detailed structure. The term is loosely used, often interchangably with the term gum. Chemically, mucilage is closely allied to gums and pectins but differs in certain physical properties. Although gums swell in water to form sticky, colloidal dispersions and pectins gelatinize in water, mucilages form slippery, aqueous colloidal dispersions which are optically active and can be hydrolyzed and fermented. Mucilages are formed in normal plant growth within the plant by mucilage-secreting hairs, sacs, and canals, but they are not found on the surface as exudates as a result of bacterial or fungal action after mechanical injury as are gums. Mucilages occur in nearly all classes of plants in various parts of the plant, usually in relatively small percentages, and are not infrequently associated with other substances such as tannins. The most common sources are the root, bark, and seed, but they are also found in the flower, leaf, and cell wall. Any biological functions within the plant are unknown, but they may be considered to aid in water storage, decrease diffusion in aquatic plants, aid in seed dispersal and germination, act as a membrane thickener, and act as a food reserve. Mucilages are commonly identified by physical properties, most recently by infrared spectroscopy.

Most mucilages are considered to be cellulosic polysaccharides containing the same group of sugars as gums and pectins, and commonly occur as salts of acids (acid type). The cations are chiefly those of calcium, magnesium, potassium, and sodium. The most common acids are uronic acids, the most important being galacturonic acid, although other acid residues including sulfonic acids are known. Some are known to exist with the acid function absent (neutral type) and consist only of simple sugars connected by a glycosidic linkage. Hydrolysis of mucilages yields pentoses and hexoses, the most common of which are arabinose, galactose, glucose, mannose, rhamnose, and xylose.

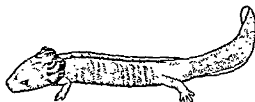
The chief industrial sources of mucilages are Icelandic and Irish moss, linseed, locust bean, slippery elm bark, and quince seed. They are obtained by milling if they occur in the endosperm, but more commonly by extraction with water or dilute sodium carbonate solution when they are found outside the seed coating. They are purified by precipitation with alcohol or salt solutions from the aqueous solution and marketed as powders. Synthetic mucilages can be formed by decomposition reactions, as in the hydrolysis of starch to form dextrins which are mixed with gums to form adhesives. Mucilages find applications in cosmetics (hand lotions and hair sets), medicinals (laxatives and diuretics), pharmaceuticals (emulsifying agents and materials for prevention of precipitation by colloidal suspension), and industry (thickening agents, emulsion stabilizers, binders, sizing

of silks, textile printing, and paper manufacture). *See* ADHESIVE; DEXTRIN; GLUE; GUM; PECTIN. [E.H.H.]

Mud puppy

A common name properly limited to the salamanders of the genus *Necturus*. The name is commonly applied to all relatively large salamanders, as is the term water dog.

The genus *Necturus* is limited to eastern North America. The most widespread and best-known species is the common mud puppy, *N. maculosus*, found over most of the United States east of the Great Plains. This salamander is completely aquatic, and retains external gills throughout life; lungs are also present. It is tan to brown or black in ground color, and variously spotted or mottled with darker pigment. Adults are 9–12 in. long. It is rather irregularly distributed in lakes and streams throughout its range; however, it may be abundant in certain localities.



Mud puppy, Necturus maculosus. (From T. I. Storer and R. L. Usinger, General Zoology, 3d ed., 1957)

The mud puppy is of slight importance as a fish bait, and is commonly dissected for study in advanced zoology classes. Its role in the aquatic economy is not clear; it probably serves as food for larger fishes, and, in turn, eats small fishes, crustaceans, insects, and fish eggs.

Mud puppies are said to live from 20 to 25 years. Many people fear them because of the erroneous belief that they are poisonous, a belief probably based on the presence of slightly irritating mucus on their slimy skins. They are reported to be edible. See SALAMANDER. [J.D.B.]

Mud volcano

A conical landform composed of clay and silt with variable admixtures of sand and rock fragments, the whole resulting from eruption of wet mud impelled upward by fluid or gas pressure. The form of such features depends upon the ratio of mineral matter to the fluids and gases in the mobile material. More liquid mud produces [illegible] constructs mud cones with slopes occasionally as steep as 40° and diameters ranging upward to several hundred yards. With repeated eruptions such cones may be built up to heights of many hundred feet (Figs. 1 and 2).



Fig. 1. Mud volcanoes near Douglaston, New York. Photograph taken at low tide, September 12, 1903 (USGS)

Associated features include mud pits, developed where the mobile material contains excessive amounts of fluid and gas, and pitch lakes, formed where the bitumens of petroleum residues are at least as voluminous as the clay particles in the viscous mud. Mud pits range in size from mere pinholes to depressions covering several acres. Minor amounts of petroleum residues are often found in mud volcanoes and associated features, but pitch lakes are comparatively rare phenomena. *See* ASPHALT AND ASPHALTITE; PETROLEUM GEOLOGY.

Occurrence. Mud volcanoes are most commonly observed in regions of geologically recent and tectonically intense orogeny, as in Trinidad, B.W.I.; the Caucasus; Java; Burma; Romania; and Colombia; or in areas of unusually rapid sedimentation with compaction lag, as along the Gulf Coast of North America or in intermontane basins at many places throughout the world.

Dynamics. The eruptive energy responsible for the formation of mud volcanoes and associated features is that of the pressure of fluids and gases accumulated in sedimentary deposits. No igneous (pyrogenetic) activity is involved. Layers or lenses of incompetent clay or silt with at least average puddling properties are ordinarily essential. Eruptions occur when the accumulated pressure in the reservoir rocks exceeds the overburden or hydrostatic pressure. This may result from orogenetic or compactional stresses, or may be due to the re-

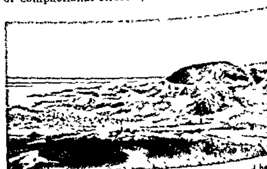


Fig. 2. Group of mud volcanoes, now submerged beneath Salton Sea, 6 miles northwest of Imperial Junction, California. (USGS)

moval of critical amounts of overburden by erosion. Compaction may be abruptly expedited by earthquake vibrations; fluid and gas pressures may be increased or decreased by various geochemical or geophysical processes.

The rate, intensity, frequency, and duration of eruptions depend on the accumulated stress potential, on the permeability and saturation of the source material, and on the physical properties of the clay or silt under stress. Activity is commonly intermittent over a period of years, but this is by no means always the case. Some occurrences, such as those on the island of Trinidad, are known to have been active since late Pleistocene time.

[H.H.SU.]

Bibliography: A. I. Levorsen, *Geology of Petroleum*, 1954.

Muffler

A device used to attenuate sound being propagated in conjunction with a moving stream of fluid (usually gas). Mufflers may be placed in two general classifications, depending upon the mode by which sound energy is removed from the gas stream. Those devices which attenuate by reflecting sound back to the source are called nondissipative, or reactive. Devices which absorb sound energy as the gas passes through them are called dissipative devices.

Selection of a particular muffler for any given application should be based upon at least the following considerations: (1) attenuation as a function of frequency; (2) length and cross-sectional area; (3) effect of gas flow on erosion, pressure drop, temperature limitations of materials, corrosion, and self-noise generated in the muffler itself during flow; and (4) initial and replacement cost. The actual acoustical performance is specified in terms of the attenuation as a function of frequency and length. The remaining considerations affect the gas flow properties and the economics of installation.

Reactive mufflers. These are predominantly used in the intake and exhaust of automotive and other reciprocating engines and compressors. Reflection is provided through the use of acoustic filters, resonators, and change in direction by bends in the pipe confining the gas stream (*see* FILTER, ACOUSTIC; RESONATOR, ACOUSTIC). These mufflers are particularly useful for low-frequency applications and for those installations where high temperatures or flammable gases restrict the use of dissipative materials.

In most cases, the design of reactive mufflers assumes the lumped parameter restriction usually applied to acoustic filters; that is, all dimensions of the device must be small as compared to a wavelength of sound. At high frequencies, where this condition is not fulfilled, the performance of the muffler cannot be predicted accurately without resorting to wave analytical procedures. Ordinarily this approach is too complicated mathematically; therefore experimental measurements are employed to determine the muffler performance.

One of the primary characteristics of a reactive muffler is its relatively high pressure drop for a given value of gas flow velocity. This pressure drop can exhibit itself, for example, as a back-pressure at the exhaust of an engine, thus restricting the engine performance. Typical back-pressures produced by passenger car mufflers range as high as 3–7 lb/in.^{2} at maximum engine power.

Examples of current practice in reactive muffler design applied to automobile exhausts are shown in Fig. 1. The "straight-through" muffler in Fig. 1*a* uses side-branch resonators to provide attenuation. The other two mufflers use combinations of resonators, filters, and flow-reversing bends to provide adequate noise control.

Dissipative mufflers. These are used in a wide range of applications where low pressure drop and high attenuation at predominantly middle and high frequencies are required. Many different configurations and types of materials can be employed, depending upon the individual application. In a simple ventilating system for a home or office the structure might consist of lining the interior walls of the duct with a blanket of glass fiber material of the order of 1 in. in thickness. On the other hand, a large wind tunnel or jet-engine test cell exhaust might require a complicated design of ducts, absorbent baffles in a parallel stack, cylinders of absorbing material mounted as a spatial array, or one of a number of proprietary designs combining these elements.

The most common dissipative muffler is obtained by lining the interior of a duct with sound-absorbing material. The approximate attenuation per lineal foot of duct provided in such a configuration can be estimated from the equation

$$\text{Attenuation per foot} = 12.6\alpha^{1.4}P/A \quad \text{decibels}$$

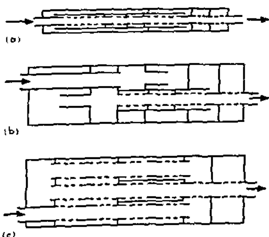


Fig. 1. Cross sections of typical reactive mufflers. (*a*) Straight-through type. (*b*) Two-tube, three-pass type. (*c*) Three-tube, three-pass type.

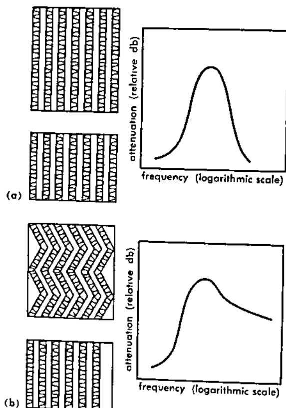


Fig 2. Plan and section views of parallel baffle mufflers. (a) Plain-parallel type. (b) Staggered-parallel type

where α is the average acoustical absorption coefficient of the lining material, P is the duct perimeter in inches, and A is the cross-sectional area of the duct in square inches. The absorption coefficient α is a function of frequency and may be obtained from material manufacturers' data. The preceding equation is most useful at low frequencies. At higher frequencies, greater values of attenuation are often achieved but cannot be predicted without consideration of the higher order wave propagation characteristics of the duct. In general, the maximum efficiency of a lined duct occurs at a frequency where the width of the duct is between about one-half and twice the wavelength of sound

Parallel baffles. These are often used for mufflers in cases where maximum efficiency requires a certain maximum spacing between the two faces of an absorbing material, but the air-flow requirements indicate that the duct should have a large cross-sectional area Parallel baffles may be looked upon as a series of ducts placed side by side in which the duct cross section is a narrow but long rectangle. The attenuation per lineal foot of a parallel baffle structure depends upon the thickness of the individual baffles and their on-center spacing. Schematic plan and section views of a set of parallel baffles and an indication of the shape of the attenuation curve are shown in Fig. 2*a*.

A variation on the parallel baffle arrangement is obtained by staggering sections of baffles in a zig zag manner, as indicated in Fig. 2*b*. The attenuation of high frequencies, that is, where the wave length is small as compared to the thickness and on-center spacing of the baffles, is considerably improved over the completely parallel configuration indicated in Fig. 2*a*.

The lined duct and parallel baffle structures are basic in dissipative muffler design. Many modifications of these principles can be performed to satisfy a particular application. Combinations of these structures with resonators have been used in developing commercial mufflers having wide frequency-range attenuation characteristics. *See* ABSORPTION (SOUND); NOISE CONTROL. [W.J.C.]

Bibliography: I. Dyer, *Noise Control*, 2(3):50, 1956; C. M. Harris (ed.), *Handbook of Noise Control*, 1957.

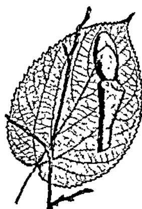
Mulberry

A genus, *Morus*, of trees characterized by milky sap and simple, often lobed, alternate leaves. Red mulberry, *M. rubra*, attains a height of 60 ft, and grows in the eastern half of the United States and southern Ontario. Its leaves are rough above, soft and pubescent (hairy) beneath, and sometimes have two or three lobes. The fruits are about 1 in. long, dark purple or black, sweet, and used as food by both domestic and wild animals. The wood is used for fence posts, furniture, interior finish, agricultural implements, and barrels.



White mulberry, *Morus alba*. (A. H. Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

White mulberry, *M. alba*, was introduced into the United States from China during the nineteenth century as a source of food for silkworms. The silkworm project was unsuccessful, but the trees remained and now are common in cities and on the borders of forests. This species has smooth, shiny

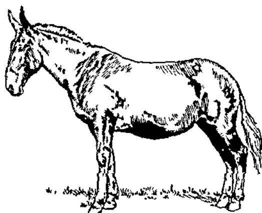


Red mulberry, *M. rubra*. (A. H. Graves, *Illustrated Guide to Trees and Shrubs*, Harper, 1956)

leaves that are usually lobed, and a very irregular habit of branching. The fruit is white and insipid. *See* FOREST AND FORESTRY; TREE. [A.H.G.]

Mule

A hybrid animal, usually sterile, resulting from the cross of the jack, or male donkey, *Equus asinus*, with the mare, or female horse, *E. caballus*. The cantankerous disposition and braying voice of the mule are well known; both are characteristics acquired from its sire. The mule is also well



The mule, *Equus asinus* × *E. caballus*; height to 64 in. (From E. L. Palmer, *Fieldbook of Natural History*, McGraw-Hill, 1949)

known for its physical toughness, exhibiting a durability under adverse conditions equalled by few other animals. Once America's most important work animal, the mule is rapidly disappearing as machines replace animals in the economy. *See* DONKEY; MULE PRODUCTION. [J.D.B.]

Mule production

The number of mules on American farms reached a peak of 5,900,000 head in 1925. Since World War II, the number has decreased about 10% annually to a total of approximately 1,000,000 in 1958. Increased mechanization of agricultural operations in the South would seem to indicate further reduction in the number of mules in future years.

The mule is a hybrid sired by a male ass (*Equus asinus*) out of a female horse (*E. caballus*). The opposite cross, very seldom made, produces the hinny (a hybrid between a stallion and a female ass). The mule and the hinny are usually sterile, but two authenticated cases of hinnies producing living progeny are known. Male mules, often called horse mules, are almost always castrated to make them more tractable as work animals.

Jacks and jennets are the males and females, respectively, of the domestic ass family. In the United States, most jacks have been used to sire mules, whereas almost all jennets have been kept solely to produce more jackstock. Small asses, about the size of Shetland ponies, are commonly called donkeys or burros (*see* HORSE PRODUCTION). They are used as pack animals in a few mountainous areas of the Southwest, and occasionally as children's pets in other parts of the country.

American jacks, sometimes called Mammoth jacks, were developed from a blending of several European strains of jackstock, principally the Catalonian, Maltese, Majorcan, Andalusian, Italian, and Poitou. Large jacks may measure over 15 hands (60 in.) and weigh 1100 lb. Compared to the horse, jacks have shorter hair on tail and mane, much longer ears, smaller and deeper hoofs, and no callosities (hard or thickened areas) on the inside of the hind legs. The loud, harsh voice is called a bray. The mule also shows these characteristics, but usually is more full bodied and more smoothly constructed than the jack.

A jack to be used for mule production should be raised with horses after weaning and not permitted to see jennets and mules until he has been broken to breed mares. Otherwise, the jack may be very slow or even unwilling to breed a mare.

Heavy draft mares bred to jacks produce big draft mules. Mares of finer quality produce lighter, finer-boned mules of the kind that were formerly used by many tobacco growers and cotton farmers in the South.

Mules are noted for their endurance, sure-footedness, and ability to stand hard work in hot weather. They can safely be self-fed in lots or corrals, whereas horses cannot. Usually steady and free from nervous excitability, mules can be handled by inexperienced or careless farm labor. *See* AGRICULTURAL SCIENCE (ANIMAL). [J.M.K.]

Multimeter

A common term for a volt-ohm-milliammeter, also called an analyzer or circuit analyzer (*see* VOLT-OHM-MILLIAMMETER). A much less common usage applies to a self-contained test instrument containing two or more single- or multiscale indicating instruments for measuring simultaneously two or more electrical quantities. [I.F.K.]

Multiple proportions, law of

This law states that when two elements combine together to form more than one compound, the weights of one element that unite with a given weight of the other are in the ratio of small whole numbers. The law can be illustrated by the composition of the five oxides of nitrogen. One gram (g) of nitrogen is combined with 2.85 g of oxygen in nitrogen pentoxide, N_2O_5; with 2.28 g in nitrogen dioxide, NO_2; with 1.71 g in nitrogen trioxide, N_2O_3; with 1.14 g in nitric oxide, NO; and with 0.57 g in nitrous oxide, N_2O. These numbers are in the simple ratio of 5:4:3:2:1. *See* DEFINITE COMPOSITION, LAW OF. [T.C.W.]

Multiple sclerosis

A fairly common degenerative nervous system disease, principally affecting young and middle-aged adults. The cause is unknown but the actual lesions are characteristic. They consist of small areas of loss of myelin, the white fatlike substance forming a sheath which surrounds many nerve fibers. There is a predilection for the optic nerves, portions of the brain stem, and the long sensory tracts of the cord, but the lesions may be widely scattered and irregular. *See* NERVOUS SYSTEM.

The disease is marked by remissions and relapses, and the rate of degeneration varies tremendously. This causes a wide range of clinical findings, extending from relatively mild temporary nerve involvement to rapid deterioration and death within a few months.

Early symptoms are abnormal sensations (paresthesias), double vision (diplopia), mild sensory and motor involvement, and cerebellar incoordination. If the disease is progressive or recurs in severe form, there is less and less normal nerve tissue that remains functional. Frequently the terminal stage is marked by dementia, blindness, and paralysis or ataxia. However, this drastic situation is not reached by all patients since arrest or remission is fairly frequent.

Recent investigations on multiple sclerosis, or disseminated sclerosis as it is sometimes called, have led to some encouraging findings concerning the causes, prognosis, and treatment. Multiple sclerosis is one of the most frequent nonvascular causes of neurologic disturbance in the 20- to 40-year age group, but it also occurs in both younger and older persons. [E.C.ST.]

Multiplexing

The means or result of the means whereby a number of independent messages, each unambiguously represented by information-bearing signals, may be sent and received over a common transmitting medium. Viewed broadly, a *multiplex* system is any part of a communication system in which two or more distinct signal channels are combined. The reason for multiplexing is either economy or necessity. Multiplexing is generic to most present-day systems of telecommunications. *See* TRANSMISSION THEORY AND METHODS.

Theoretically, many types of multiplexing are possible. Typical examples are time division, frequency division, and phase discrimination. In time division, messages are multiplexed by sequentially allocating different time intervals for the transmission of each message; in frequency division, different frequency bands are allocated to different message channels; and in phase discrimination there need be no separation of channels in either frequency or time, that is, each message channel may utilize the entire available frequency spectrum all of the time. *See* AMPLITUDE MODULATION. [H.S.BL.]

Multiplication

One of the fundamental operations of arithmetic and algebra. The use of the symbol ×, commonly employed in arithmetic to denote multiplication, is attributed to the English mathematician William Oughtred (1574–1660). Because of its resemblance to the letter x, it is rarely used in algebra, where multiplication is frequently denoted by a dot (as in $a \cdot b$) or, most often, merely by juxtaposition of letters (for example, ab). Multiplication of numbers (real or complex) is associative, $a(bc) = (ab)c$; commutative, $ab = ba$; and distributive with respect to addition, $a(b + c) = ab + ac$; but the term has been extended to denote binary operations on many other kinds of objects, and these operations need not possess all the properties of ordinary multiplication listed above (for example, multiplication of matrices is not commutative). Much effort was formerly expended in the computation and design of multiplication tables, and the related endeavor of tables of prime numbers. (A whole number is a prime if it is not exactly divisible by any whole number other than itself and 1.) *Crelle's Journal* (1895) gave a 1000 × 1000 table, and D. H. Lehmer has given a table of all primes from 1 to 10,006,721. Electronic calculators are well adapted for such work and modern investigations in that area are now done almost exclusively with their aid. *See* ADDITION; ALGEBRA; DIVISION; NUMBER SYSTEMS; NUMBER THEORY; SUBTRACTION. [L.M.BL.]

Multipole radiation

Gamma rays, internal conversion electrons, or positron-electron pairs of defined characteristics emitted from an atom when the nucleus makes a transition between two energy states. The *multipole order* is the number of units of angular momentum removed by the radiation. This number is not necessarily equal to the difference between the spins of the nucleus in its initial and final states, because the nuclear spin direction may change. Thus a quadrupole radiation will result when a state of spin 2 makes a transition to a state of spin 0, but a transition from a state of spin 2 to one of spin 1 may also result in quadrupole radiation if

Spin and parity changes for various multipoles

ΔI	0	1	2	3	4	5
Multipole	Monopole	Dipole	Quadrupole	Octupole	2^4-pole	2^5-pole
Electric	E0 no	E1 yes	E2 no	E3 yes	E4 no	E5 yes
Magnetic		M1 no	M2 yes	M3 no	M4 yes	M5 no

there is an appropriate change in nuclear spin direction. If I_i and I_f are the spins of the nucleus in its initial and final states, then the multipole order ΔI must be such that

$$|I_i - I_f| \leqq \Delta I \leqq |I_i + I_f|$$

in order to conserve angular momentum.

From multipole radiation measurements, the static and dynamic properties of nuclear energy states may be determined, and this information may be used in theories of nuclear structure.

In addition to the energy and angular momentum values, a third characteristic of a nuclear state is the parity. The parity of a given state is odd (−) or even (+) depending on whether the quantum-mechanical wave function describing the nucleus in that state changes sign upon transposing the function to a reflected coordinate system. There are two classes of multipole radiation, the electric and the magnetic, and the designation of a given radiation depends upon both the angular momentum change and whether the parities of the initial and final states of the nucleus are the same. The table illustrates the spin and parity changes for the various multipoles. See PARITY (QUANTUM MECHANICS); SPIN (QUANTUM MECHANICS).

The characteristics of multipole radiations may be determined experimentally by measurements of one or more of the following: the angular correlation of the γ-ray with another coincident or cascade γ-ray; the relative intensities of internal conversion electrons and γ-rays; the relative intensities of internal conversion electrons from various electron shells of the atom; the characteristics of the internal positron-electron pair formation; and the half-life of the transition. Nuclear transitions generally take place with emission of the lowest possible order of multipole radiation, because the higher the order the longer will be the half-life. See RADIOACTIVITY; SELECTION RULES (PHYSICS).

[D.F.A.]

Bibliography: J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics*, 1952.

Multituberculata

The small, probably quadrupedal, multituberculates were the only Mesozoic mammalian herbivores. They were abundant in Late Jurassic, Late Cretaceous, and early Paleocene faunas of North America and Eurasia but toward the end of the Paleocene they became quite rare and apparently were extinct by the end of early Eocene time. The morphologic specializations of the multituberculates foreshadow many of the specializations of modern rodents. The appearance and early diversification of rodents in late Paleocene and early Eocene time coincided with the final decline and extinction of the multituberculates. This suggests that competition between the two groups may have been a significant factor in the extinction of the multituberculates. See RODENTIA

Multituberculates had two upper and two lower molars with cusps arranged in two or three longitudinal rows (Figs. 1 and 2). The premolars usually functioned as shearing teeth. Upper and lower incisors tended to be reduced to two rodentlike teeth. The mandible did not have a distinct angular process.

The Late Jurassic and Middle Cretaceous plagiaulacids had three or four bladelike, lower premolars with serrate crests and ribbed sides. These teeth sheared against the five upper premolars which, though smaller and with fewer cusps, resembled the lower molars rather than the lower premolars. The upper and lower molars had two longitudinal rows of from two to four cusps each.

The Plagiaulacidae probably gave rise to the Ptilodontidae which are known to have lived from Late Cretaceous to early Eocene time. The lower premolars of the ptilodonts were reduced to a single, large, bladelike tooth usually preceded by a small, peglike tooth. The number of molar cusps of

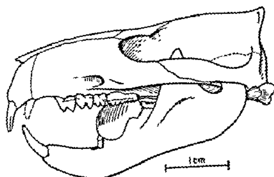


Fig. 1. Skull and jaw of *Ptilodus*, an early Tertiary multituberculate. (After G. G. Simpson, 1937)



Fig. 2. Occlusal view of a first upper molar of a multituberculate.

ptilodonts was increased and a third row of cusps was added to the upper molars. There were, at most, four upper premolars. In some ptilodonts (Ptilodontinae) the enamel encircled the crown of the lower incisors; while in others (Eucosmodontinae) the enamel was restricted to a thin anterior band.

Members of the Paleocene Taeniolabidae had only single, peglike upper and lower premolars. Their molars were enlarged and the number of molar cusps was increased. *Taeniolabis*, the largest known multituberculate, was no larger than a small terrier.

The ancestry of the multituberculates is unknown. Though some workers have speculated that rodents or other groups might be closely related to the multituberculates, it is now fairly certain that the multituberculates were not ancestral to any other order of mammals. See ALLOTHERIA.

[W.A.CL.]

Multivibrator

A form of relaxation oscillator consisting normally of two active devices, such as vacuum tubes or transistors, interconnected by electric networks. A portion of the output voltage of each is applied to the input of the other and is of such magnitude and polarity as to maintain the devices alternately conducting over controllable periods of time. The transition time of each device from one state to the other is extremely short. The voltage waveform from the output of each of the devices is essentially rectangular in form.

There are various classifications of multivibrators, which are based on the manner in which the reversal-of-state action of each device is initiated and upon the method of control of the time interval in each state, whether from external sources or by the decay of voltage across a capacitor in circuits containing *RC* time constants within the multivibrator itself.

Bistable multivibrator. In bistable multivibrators either of the two devices may remain conducting, with the other nonconducting, until the application of an external pulse. Such a multivibrator is said to have two stable states. A basic form of vacuum-tube bistable multivibrator along with voltage waveforms associated with one of the tubes is shown in Fig. 1. This circuit is known as the Eccles-Jordan multivibrator (after its inventors) and is commonly called a "flip-flop." The resistance networks between the positive and negative supply voltages are such that, with no current flowing to the plate of one tube, the voltage at the grid of the second is slightly negative, zero, or limited to a slightly positive value. The resultant current in the plate circuit of the second tube causes a voltage drop across the plate load resistor, which in turn lowers the voltage at the grid of the first tube to a sufficiently negative value to continue to prevent plate current flow. This condition of one tube OFF and the second ON will be maintained as long as the circuit remains undisturbed.

If a fast negative pulse is then applied to the grid of the ON tube, its plate current decreases, and its plate voltage rises. A fraction of this rise is applied to the grid of the OFF tube and causes some plate current to flow. The resultant plate voltage drop, transferred to the grid of the ON tube, causes a further rise at its plate. The action is thus one of positive feedback, with the cumulative result be-

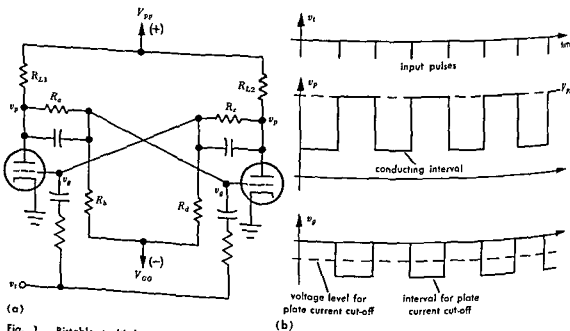


Fig. 1. Bistable multivibrator. (a) Circuit diagram. (b) Waveforms for one tube. Other tube would have identically shaped, but oppositely phased, waveforms.

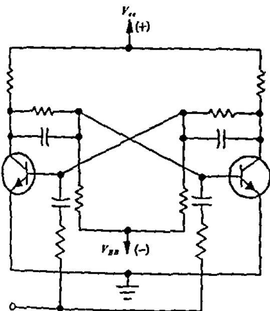


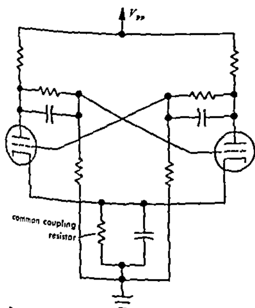
Fig. 2. Transistor bistable multivibrator.

ing a nearly instantaneous transfer of current flow from one tube to the other. There is one such reversal each time a pulse is applied to the grid of the ON tube. Normally, pulses are applied to both tubes simultaneously so that whichever tube is ON will be turned off by the action. The capacitances between the grid of one tube and the plate of the other play no role other than to improve the high-frequency response of the voltage divider network, by compensating for the input capacitances of the tubes, and thereby improving the speed of the transitions.

A transistor counterpart of the vacuum-tube bistable multivibrator with *n-p-n* transistors is shown in Fig. 2. The base of the transistor corresponds to the grid, the emitter to the cathode, and the collector to the plate. Although the waveforms are of the same polarity and the action is roughly similar to that of the vacuum-tube circuit, there are some important differences. The effective resistance of the base-emitter circuit, when it is forward biased and being used to control collector current, is much lower than the input grid resistance of the vacuum tube when used to control plate current (a few thousand ohms compared to a few megohms). This fact must be taken into account when designing the divider networks. Normal operation of the transistor circuits is achieved with much smaller terminal voltages than those of vacuum-tube circuits, and supply voltages are correspondingly lower.

For the circuits described, the waveform amplitudes are usually limited by saturation. For the vacuum-tube circuit, limiting is the usual grid circuit limiting of the tube, when the grid becomes positive with respect to the cathode and the grid input circuit has a vastly lowered impedance. For the transistor circuit, the input impedance is always low for any level of collector conduction and is not responsible for limiting. Limiting occurs when the collector voltage becomes so low that the collector-emitter voltage is lower than the emitter-base voltage. At this point, the collector becomes forward biased and the collector impedance becomes very low. When this occurs, sufficient positive feedback is not available to drive the outputs to greater amplitudes.

Certain limitations in the speed of transition between states in transistor multivibrators occur when saturation takes place. Therefore nonsaturating circuits are often employed. These involve the



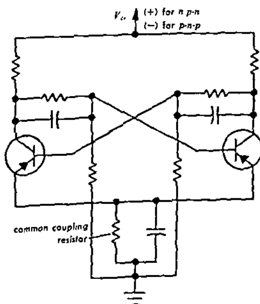
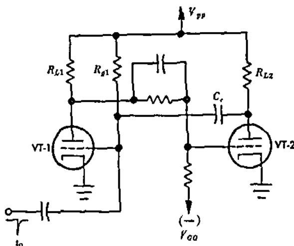


Fig. 3. Cathode and emitter-coupled bistable multivibrators.



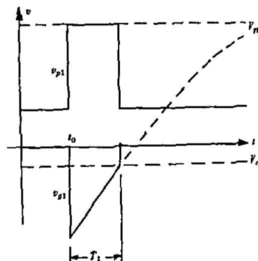


Fig. 4. Basic monostable multivibrator.

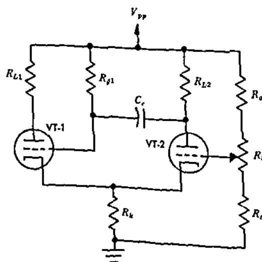
use of diodes in the coupling networks to limit the voltage levels before the transistors are allowed to saturate.

One of the voltage supplies in the bistable multivibrator often is eliminated by the use of the common coupling resistance shown in the circuits of Fig. 3. Since one of the devices is always conducting, the cathode or emitter is maintained at a constant potential. This potential makes it possible to design a divider network to maintain the desired operating levels.

The trigger pulses applied to a bistable multivibrator need not occur at a constant rate, but if they do, the output waveform will be recurrent at a constant rate having a total period equal to twice the period of the input pulses. In any case, either output goes through a complete cycle and reverts to its initial state for every two input pulses. For this reason it is often referred to as a scale-of-two or a binary circuit. Such bistable circuits play an important role in the arithmetic operations of digital computers, in modern frequency- and time-measuring equipment, and in various combinations of circuitry in many pulse systems.

Monostable multivibrator. A monostable multivibrator has only one stable state. If one of the normally active devices is in the conducting state, it will remain so until an external pulse is applied to make it nonconducting. The second device is thus made conducting (and the first nonconducting) and it will remain so for a time duration dependent upon *RC* time constants within the circuit itself.

A typical vacuum-tube monostable multivibrator is shown in Fig. 4. The input of VT-1 is capacitance coupled to the output of VT-2. In the absence of external pulses, VT-1 is conducting, with its grid at zero potential, or limited to a slight positive value by saturation. The resultant plate current limits the plate voltage to a value that makes



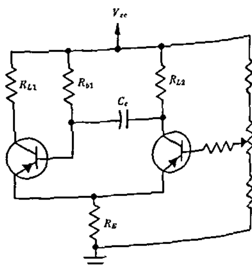


Fig. 5. Simplified monostable multivibrator.

the grid of VT-2 sufficiently negative to keep VT-2 cut off. If a negative pulse is applied to the grid of VT-1, VT-1 is cut off, and VT-2 conducts. The circuit action is similar to that of the bistable circuit, except that the voltage drop at the plate of VT-2 is transferred to the grid of VT-1 through the capacitance C_c. This change, or transition, cannot be maintained indefinitely, because the current flowing through C_c and R_{g1} causes a decrease in voltage with time across C_c and rise in voltage at the grid of VT-1, as shown. The initial drop is of the same magnitude as that at the plate of VT-2 at the time the trigger is applied. The ensuing rise is exponential in form and (if R_{g1} is much greater than R_{L2}) is given by the equation

$$v_{g1} = (V_{g1} - V_{pp})e^{-(t/R_{g1}C_c)} + V_{pp} \qquad (1)$$

When the rising voltage reaches the level V_{c1}, plate current again flows in VT-1. The positive feedback quickly causes VT-1 to become fully conducting and limited by grid-current saturation. The duration of the nonconducting interval for VT-1 (T_1 in Fig. 4) is found by solving Eq. (1) for $t = T_1$ and for $v_{g1} = V_{c1}$. The result is

$$T_1 = -R_{g1}C_c \ln \frac{V_{c1} - V_{pp}}{V_{g1} - V_{pp}} \qquad (2)$$

The circuit will remain in this state until another initiating trigger is applied.

As in the bistable circuit, transistors may be used instead of vacuum tubes, with the design appropriately modified to take the difference in their characteristics into account. Also, as in the bistable circuit, *n-p-n* or *p-n-p* transistors may be used, depending upon the desired polarity of the output waveforms. One of the supply voltages may be eliminated by using a common cathode or emitter resistance in the same manner as for the bistable circuit. In slightly different form, the divider network itself may be eliminated by using the cathode or emitter resistor as a coupling element. Such a cathode-coupled monostable multivibrator is shown in Fig. 5. Normally, VT-1 is in grid-current saturation, and its plate current determines the potential at the common cathode terminals. The voltage at the grid of VT-2 is adjusted so that its plate current is cut off under this condition. If a negative pulse is applied to the grid of VT-1, the resultant decrease in plate current lowers the cathode potential and plate current flows in VT-2, lowering the voltage at the grid of VT-1 still more. Thus the positive feedback action (causing the rapid reversal of state) is through the common cathode resistance. The duration T_1 of the OFF period of VT-1 is determined partly by the $R_{g1}C_c$ time constant in the same manner as it is for the circuit of Fig. 3, and is controllable by the setting of R_b. This setting determines how heavily VT-2 conducts and therefore controls the amplitude of the transition V_{g1} resulting from this plate current flow.

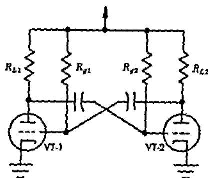


Fig. 6. A stable (free-running) multivibrator.

The monostable multivibrator has frequent application in electronic circuits where timing and gating operations are required. In particular, if a pulse or "gate" is required to be an accurate function of a control voltage, the circuit of Fig. 5 is particularly applicable. The duration of the gate interval T_1 can be made linearly proportional to the dc voltage applied to the grid of VT-2.

Astable multivibrator. The astable multivibrator has capacitance coupling between both of the two active devices and therefore has no permanently stable state. Each of the two devices functions in a manner similar to that of the capacitance-coupled half of the monostable multivibrator as shown in Fig. 6. It will therefore generate a periodic rectangular waveform at the output with a period equal to the sum of the OFF periods of the two tubes. The duration of each of the two periods is governed by an equation of the form of Eq. (2), with the appropriate values for each of the two parts of the circuit used. The transistor astable multivibrator similarly functions as the combination of two transistor monostable sections coupled together.

Astable multivibrators, although normally free-running, can be synchronized with input pulses recurrent at a rate slightly faster than the natural recurrence rate of the device itself. This is illus-

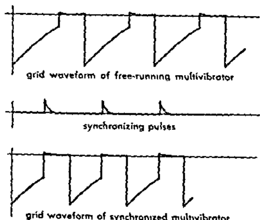


Fig. 7. Waveforms of synchronized multivibrator.

trated in Fig. 7, which shows the relation between the grid waveform and applied synchronizing pulses. If the synchronizing pulses are of sufficient amplitude, they will bring the grid to the conduction level at an earlier than normal time and will determine the new recurrent rate.

Triggering of multivibrators. The period of the bistable, the monostable, and the synchronized astable multivibrator is controlled by pulses (triggers) from an external source. These triggers may be applied to the circuit in various fashions. The initiating trigger pulsewidth should be sufficiently wide that the circuit can respond (as limited by its high-frequency response) before the pulse is over, but not so wide as to interfere with the normal action of the multivibrator once the transition has taken place. It should be coupled to the multivibrator through a small capacitance so that the loading effect of the trigger source is negligible. Usually, a faster transition can be achieved if the triggers are used to turn off a normally ON device. Triggers are usually applied to the appropriate grid, but they may also be applied to the plates and reach the grid through the coupling networks. In some cases, any coupling between the trigger source and the multivibrator is objectionable, and an isolating amplifier is used, with its plate and the plate of the multivibrator connected together. [G.M.C.]

Bibliography: B. Chance et al. (eds.), *Waveforms*, 1949; J. Millman and H. Taub, *Pulse and Digital Circuits*, 1956.

Mu metal

The name that was given originally to the ferromagnetic alloy containing approximately 78% nickel, 5% copper, and 17% iron. Now chromium and sometimes molybdenum additions to the basic iron-nickel alloy are made.

Mu metal is of value primarily for its high initial permeability, about 20,000 gauss/oersted. It can be used in the construction of transformers for weak signals, for magnetic shields, for magnetic loading, and for various electromechanical devices.

For composition and properties, *see* MAGNETIC MATERIALS. [R.M.BO.]

Mumps

An acute contagious viral disease, characterized chiefly by enlargement of the parotid glands. *See* ANIMAL VIRUS.

The virus is 175–200 mμ in diameter; it will grow in monkeys, newborn hamsters, embryonated eggs, and tissue cultures. *See* MYXOVIRUS.

Besides fever, the chief signs and symptoms are the direct mechanical effect of swelling of glands or organs where the virus localizes. One or both parotids may swell rapidly, producing severe pain when the mouth is opened. In orchitis, the testicle is inflamed but is enclosed by an inelastic membrane and cannot swell; pressure necrosis produces atrophy, and if both testicles are affected, sterility may result. The ovary may enlarge, without sequelae. Meningoencephalitis and thyroiditis are seen, but rarely. Cases without parotitis occur; 30–40% of cases are subclinical.

Diagnosis is by isolation of virus from acute saliva or cerebrospinal fluid in chick embryos or monkey kidney tissue cultures; or by complement fixing or neutralizing antibody responses. *See* CULTURE, EMBRYONATED EGG; CULTURE, TISSUE.

Mumps is endemic throughout the world, in all seasons. Epidemics occur under crowded conditions, particularly among children 5–15 years old, but also among troops. The only known reservoir is man. In cities, a high proportion of the population is immune by the age of 15 years.

Gamma globulin prepared from mumps convalescent serum, given after parotitis has appeared, can decrease the incidence of orchitis. A killed virus vaccine is available, but because mumps is usually mild, widespread immunization has not been recommended except for adults at special risk. *See* BIOLOGICALS; GAMMA GLOBULIN. [J.L.M.]

Bibliography: T. M. Rivers and F. L. Horsfall, Jr. (eds.), *Viral and Rickettsial Infections of Man*, 3d ed., 1959.

Muonium

An "atom" consisting of an electron bound to a positively charged muon (μ-meson) by their Coulomb attraction, just as an electron is bound to a proton in the hydrogen atom. The pattern of energy levels for muonium should closely resemble that for hydrogen, with small differences due to the smaller mass and larger magnetic moment of the muon. Because of the inherent instability of the muon, even isolated muonium would live for only about 2.2 microseconds. In coming to rest in matter of low density, positive muons are expected to capture atomic electrons and form muonium. By 1960, however, no direct evidence had been obtained for the formation of muonium. *See* MESON; POSITRONIUM. [R.H.D.]

Musci

A class of small green plants commonly called mosses which make up the largest and most highly developed group of the phylum *Bryophyta*. They are world-wide in distribution, occurring in nearly all damp habitats except the ocean. They are abundant on soil, rocks, and other plants in moist woodlands and may form conspicuous carpets of vegetation in arctic and alpine regions.

Structure. The moss plant consists of a prostrate or erect stemlike structure bearing leaflike appendages. Internally, certain cells differentiate, forming elongated conducting strands, but no true vascular system is present as in higher plants (*see* PHLOEM; XYLEM). The plants are anchored to the substratum by septate, usually branched, rootlike structures called rhizoids. The green stalk and leaflike structures contain chlorophyll and carry on photosyn-

thesis. Therefore, these plants are independently self-sustaining (*see* CHLOROPHYLL; PHOTOSYNTHESIS).

Alternation of generations. The green gametophyte bears the sex organs, antheridia (male) and archegonia (female). These are found in the tips of the stemlike structures, either on separate plants, on separate branches of the same plant, or immersed in the "leafy" apex of a single branch. The sperms, or male gametes (produced in the antheridia) are biciliate or biflagellate and motile

the sporangium, each spore may develop into a filamentous or thalloid protonema which produces lateral buds that grow into new gametophytes.

Classification. Mosses have been divided into three groups: the Sphagnopsida, or peat mosses; the Mnionopsida, or true mosses; and the Andreaeopsida, or granite mosses.

The peat mosses contain only a single genus, *Sphagnum*, which nevertheless has hundreds of species (Fig. 1). In size these are the largest of mosses, being perennial and with practically indefinite growth (*see* PERENNIAL PLANTS). They are markedly social, forming extensive masses on boggy

suing fragmentation accounts for the dense cushions of closely packed shoots. *Sphagnum* differs from other mosses in having a broadly thallose protonema which produces but one gametophore, or upright branch. The leaflike structures are without a midrib and are composed of two kinds of cells. One type soon dies, resulting in cells which are

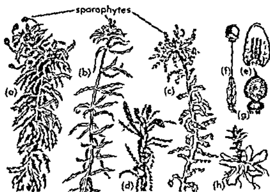


Fig. 1. (a, b, c) Three species of *Sphagnum*. The small urn-shaped bodies are sporophytes. (d, e, f, g) Reproductive structures; (g) represents the sporophyte stage, which consists of a sporangium and a very short stalk. (h) Thalloid protonema with young leafy shoot. (*From E. N. Transeau, H. C. Sampson, and L. H. Tiffany, Textbook of Botany, rev. ed., Harper, 1953*)

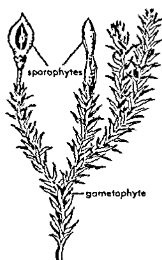


Fig. 2. A granite moss, *Andreaea rupestris*, showing a gametophyte with mature and immature sporophytes. (*From P. M. Smith, Cryptogamic Botany, vol. 2, 2d ed., McGraw-Hill, 1955*)

long and wide with perforated walls and internal thickenings. These are the cells which absorb and hold water. The other cells are long lived, narrow, and contain numerous chloroplasts. Both the male sex organs (*antheridia*) and the female sex organs (archegonia) are borne on short lateral branches. The sporophyte which develops from the fertilized egg (zygote) has a bulbous base (pseudopodium) embedded in the gametophyte. The seta, or stalk, which supports the spherical sporangium, or capsule, is reduced but its shortness is compensated for by the growth of the gametophyte (pseudopodium) which projects the sporangium well above the vegetative tissue. In other respects the life cycles of peat mosses are similar to those of other mosses.

Peat moss is used as a fuel, a mulch, and for packing the roots of ornamental seedlings. *Brazil* is the chief center of commercial distribution.

The granite mosses are represented by only

masses on siliceous rocks. The general structure resembles that of the true mosses, but in several details they are like the peat mosses. Their one unique feature is the longitudinal splitting of the mature capsule into four valves.

The true mosses are the largest group of the Musci and include all the common genera (Fig. 3). Despite their abundance and wide distribution, they display great uniformity of structure and have a common life history. They differ from the other Musci in having a filamentous protonema, a midrib in their leaflike structures, *no pseudopodium and no differentiation of tissue within the capsule.*

Whereas the peat mosses and the granite mosses are relatively limited in geographical distribution.

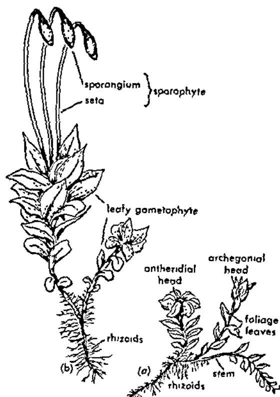


Fig 3. A true moss, *Mnium*. (a) Location of sex organs (male antheridia and female archegonia). (b) A gametophyte with attached sporophytes. (From W. W. Robbins, T. E. Weier, C. R. Stocking, *Botany: An Introduction to Plant Science, 2d ed*, Wiley, 1957)

the true mosses are found from the polar regions to the tropics. They may grow on soil or rocks, in bogs, marshes, ponds or streams, or as epiphytes on trees or leaves. Most of them live in moderately moist habitats. Many are pioneer species, colonizing any opening in the vegetation. In this respect they are of considerable importance in forestalling or stopping erosion.

The life cycle of a true moss plant starts with a single-celled, asexual reproductive structure, the spore (Fig 4). In a favorable environment, the spore germinates to produce a much-branched filament or threadlike structure called the protonema. The latter is composed of numerous cells, each containing chloroplasts. However, as more branches are produced, some of the filaments extend into the substrate and function as absorbing and anchoring structures, in which case they lose their chlorophyll and become rhizoids.

In most moss plant genera, the protonema is a transitory structure which produces buds by the division of apical cells. These buds give rise to erect leafy branches, the gametophores, which bear the sex organs of the gametophyte. These adult plants (gametophytes) develop their own anchoring rhizoids and have a stemlike axis bearing leaflike structures containing chloroplasts. These green "leaves" carry on photosynthesis and therefore these gametophytes are capable of an independent self-sustaining existence. The mature "leaves" in most species are only one cell in thickness, except at the midrib, and they are arranged in three vertical rows because the apical cell from which they develop has three faces where the divisions occur. Stomata are absent from all parts of the gametophyte. The "stem" may be somewhat differentiated into a central strand of vertically elongated cells surrounded by cortical, chloroplast-bearing cells and an outer epidermis. True vascular tissues (phloem and xylem) are absent.

The gametophores produce the sex organs (antheridia and archegonia). In the different genera, the relative position of the antheridia and archegonia varies. In some species both occur on the same "stem" apex; in others there are antheridial and archegonial branches on the same plant, and there are many species in which the antheridia and archegonia are borne on separate plants. Interspersed among the sex organs there are modified leaves called paraphyses which may function to keep the organs moist. In most cases, the sex organs can be seen without magnification. Antheridia and archegonia are borne on pedicels. Because of their extremely long, twisted necks, the archegonia are quite conspicuous. In many species, as the antheridia mature, the chloroplasts in the jacket, or outer sterile cells, become chromoplasts and give the apex of the plant a marked reddish color.

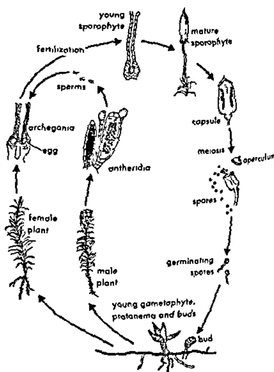


Fig. 4. The life cycle of a moss. (From E. W. Sinnott and K. S. Wilson, *Botany: Principles and Problems*, 5th ed., McGraw-Hill, 1955)

The biflagellate sperms, released only when the mature antheridium is covered with water, swim rapidly and freely. By swimming the sperms reach the archegonial areas and eventually arrive at and unite with the eggs. The resulting zygotes are retained in the archegonia where each zygote develops rapidly into a spindle-shaped embryo. As it grows, the embryo produces at its basal end a foot embedded in the gametophyte and at the other end an elongated stalk, or seta, bearing a terminal capsule, or sporangium (Fig. 5). As the capsule is produced, part of the archegonium enlarges and is converted into a cap, or calyptra, which is torn loose and carried upward on the capsule as the seta elongates. Chloroplasts and stomata are present in the outer cells of the stalk and capsule tissues of the sporophyte. This is the basis for the suggestion that the sporophytic generation may be dependent upon the gametophyte only for water and mineral elements.

When the capsule matures, a lid, or operculum, is exposed. A ring of teeth known as the peristome develops under the operculum and within the capsule. These teeth are moisture sensitive, bending outward when the air is dry and turning inward when the humidity is high, and in this way they facilitate the distribution of the spores. The spores germinate and produce the protonema, thus completing the life cycle. The dependence on water as a medium necessary for fertilization is considered a weakness that has restricted this group to amphibious habitats and is responsible for the small size of the plants. *See* PEAT; TRACHEOPHYTA.

[P.A.V.]

Bibliography: *See* BRYOPHYTA.

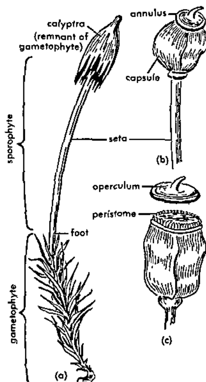


Fig. 5. Diagrams of (a) complete moss plant; (b) tip of sporophyte with calyptra removed; (c) capsule (sporangium) showing the operculum and peristome. (*From F. W. Emerson, Basic Botany, 2d ed., Blakiston, 1954*)

Muscle

The tissue of the body in which cellular contractility has become most apparent. Almost all forms of protoplasm show some degree of contractility, but in muscle fibers adaptation has led to the preeminence of this property. In vertebrates three major types of muscle are recognized: smooth, cardiac, and skeletal. *See* MUSCLE (BIOPHYSICS).

Smooth muscle. Smooth muscle, also designated as visceral and sometimes as involuntary, is the simplest type. Such a muscle consists of elongated fusiform cells which contain a central oval nucleus. The size of such fibers varies greatly, from a few microns up to 0.5 millimeter in length. These fibers contract relatively slowly and have the ability to maintain contraction for a long time. Smooth muscle forms the major contractile elements of the viscera, especially those of the respiratory and digestive tracts and the blood vessels. Smooth muscle fibers in the skin allow for regulation of heat loss from the surface. Those in the various ducts of the body act to carry or move the contents to their destinations, as in the biliary system, ureters, and reproductive tubes.

Smooth muscle is usually arranged in sheets or layers which commonly are oriented in different directions. Functionally, the major features of smooth muscle are its intrinsic ability to contract spontaneously and its dual regulation by the autonomic nerves of the sympathetic and parasympathetic divisions.

Cardiac muscle. Cardiac muscle has features common to smooth muscle in that it is innervated by the autonomic system, yet retains the ability to contract spontaneously. This rhythmic contraction begins early in embryonic development and continues until death. Variations in the rate of contraction are induced by autonomic regulation and many other local and systemic factors. *See* AUTONOMIC NERVOUS SYSTEM.

The cardiac fiber, like smooth muscle, has a central nucleus, but the cell outline is elongated and not symmetrical. The term syncytium is used to denote the branching intertwining of protoplasmic elements without definite cell boundaries.

Histologically, the features of cardiac muscle include longitudinal striations and the presence of transverse bands, called intercalated disks, which appear at short intervals.

The heart contains its own specialized system for initiation and spread of contraction in wavelike form over the myocardium. This is the conducting system, composed of the sinoauricular and atrioventricular nodes and the intervening bundles of special tissues that transmit the primary impulses.

The conducting system is modified cardiac muscle, sometimes called the neuromuscular system to indicate its dual characteristics. *See* CARDIOVASCULAR SYSTEM; HEART (GANGLIONIC PACEMAKER).

Skeletal muscle. Skeletal muscle is also called striated, somatic, and voluntary muscle, depending on whether the description is based on the appearance, the location, or the innervation. The individual cells, or fibers, are distinct and vary greatly in size. Some of the larger are said to be 15 cm or more in length, while the smallest are less than 1 mm. These fibers do not ordinarily branch and they are surrounded by a tissue membrane, the sarcolemma. Within each fiber several nuclei may be seen, thus the unit is actually a syncytium.

The transverse striations of skeletal muscle are quite characteristic. These bands and the intervening lighter portions occur with great regularity and at close intervals. Correlations between the histologic appearance of skeletal muscle and the biochemistry of contraction have been made but all the details are not yet available. *See* SPECIALIZED TISSUE.

Muscle morphology. When organized into muscles, skeletal fibers are arranged in an orderly fashion with general orientation between the two points of attachment. Surrounding each sarcolemma is a connective tissue layer, the endomysium. Groups of fibers form bundles, or fasciculi, which are ensheathed with connective tissue containing elastic and collagenous fibers. This perimysium is continuous with that of other fasciculi, and the group of bundles which form the entire muscle is covered with an epimysium. At points of attachment the connective tissue sheaths show a rather abrupt transition to the dense collagenous structure of a tendon.

The size of a muscle and of its component fibers is directly related to the function of that muscle. It is characteristic of many skeletal muscles that the muscle bundles may be arranged obliquely to the long axis. This arrangement allows for a larger number of fibers to exert contraction in a relatively small space and is therefore conducive to a powerful movement through a restricted distance.

Skeletal muscle groups tend to be distributed in relation to a movable part in a manner to produce antagonistic actions when appropriate nerve stimulation is initiated. In other cases, a complex movement could not be achieved by the contraction of any one muscle, so that such movement is achieved by synergistic contraction of appropriate groups.

The innervation of the skeletal muscles by the motor fibers of the peripheral nervous system presents a very constant pattern. This pattern develops early in embryonic life when motor nerves grow into the muscle tissue of each body segment. With the formation of limbs and specialized structures of the trunk, neck, and head, a great deal of rearrangement occurs, but the early relationships are not lost if the muscle has survived as a postembryonic structure.

Muscle activity. The muscles are generally divided into groups to designate their areas of action. Thus the muscles of the head include those of the face, of mastication, of the orbit, and so on.

The trunk muscles may be either intrinsic or extrinsic; that is, they may have both their origins and insertions on the trunk, or may have one point of attachment on a limb or the neck. The small oblique muscles between the ribs, the intercostals, represent the best example of remaining segmentation. The great muscles of the back allow powerful movement of the vertebral column or fix it in place so that the limbs may act, as on a fulcrum. In the regions of the shoulders and hips, functional muscle groups provide flexion, extension, and some degree of rotation of the arms or legs, as the case may be. The muscles of the abdominal wall are mostly supportive in nature, but are also involved in respiration when needed.

The arms and legs are supplied with muscles in a somewhat similar manner. Functional groups, usually contained within a connective tissue compartment, act in antagonistic and synergistic patterns to achieve the desired motion. In the hand where fine muscular movement and coordination are highly developed, the complexity of muscle-nerve coordination has reached a high point.

Similar patterns of muscle arrangement are found in all vertebrates, with allowances for differences in position and habitat. In fact, the constancy of the nerve-muscle relationship of embryonic development provides an excellent key to classification and comparative anatomy of vertebrates, since variations peculiar to certain classes or orders can be readily recognized. *See* MUSCULAR SYSTEM. [E.C.ST]

Muscle (biophysics)

From a biological point of view, muscle is a specialized tissue which, to a high degree, displays the property of mobility more or less inherent in cells in general. From the physical standpoint, muscle is one of the many systems in nature that can exert a force. It takes a special position among these, because the introspective notion of the tension associated with muscular contraction has been of crucial significance in the origination of the concept of "force." Depending on the anatomical or experimental arrangements, contraction can manifest itself by shortening while a smaller or greater weight is lifted; this is called isotonic contraction if the load on the muscle is constant. Or, when no shortening is permitted, the muscle displays tension in what is called an isometric contraction. Finally, matters can so be arranged that the muscle lengthens during its activity. To accomplish this, the muscle has to stretch during its activity with a force greater than the one it actively exerts, but this is not as artificial a situation as it might seem. In rhythmic body motions, that is, when a limb is being moved by one group of muscles (for example the flexors) and when this motion furthermore per-

sists because of the inertia of the moving mass, the subsequent activity of the antagonistic muscles (the extensors) will start while they are being stretched.

Muscles are not the only physical systems in which the exerted force can be switched on and off. The same would be true, for example, for an electromagnet in which the current can be closed and opened. In muscle, however, this switching is accomplished by a unique and highly characteristic biological phenomenon involving the concepts of stimulus, excitation, and irritability. A few fundamentals of this phenomenon are discussed in the following paragraphs.

Stimulus and response. A stimulus imposes upon a living cell a state of heightened reactivity or excitation which, in typical instances, has the property of propagating itself to neighboring regions of the cell, and of leading to a response typical for the nature of the substrate tissue. Propagation of such an impulse is most pronounced in cells that have one long linear dimension, such as muscle or nerve. In the former, the specific response consists of the contraction and its accompanying metabolic reactions elicited by the impulse. In the latter, there is no grossly observable response, other than the mere propagation of the state of excitation itself. This is accompanied by measurable physical attributes, notably the action potential, through which the progression of the impulse can be objectively established; and by altered metabolism and heat production indicative of the active participation of the cell. The amount of energy involved in the nerve is, however, several orders of magnitude below the energy released in the stimulation and response of muscle.

The bioelectric phenomena, described as action potential, which accompany the impulse are indicative of profound changes in the electrochemical configuration of a section of the cell membrane (*see* BIOPOTENTIALS AND ELECTROPHYSIOLOGY). One net result is a reversal of charge, the excited area *showing a negative variation with respect to the* resting sections. The currents which flow as a result of this potential difference are the direct cause of the propagation of the state of excitation into the next section of the cell. The resulting excitatory state and its reversal are also correlated with chemical events in a manner which is as yet poorly understood. These may differ in various tissues, but the fundamental principles of the electrical transmission of the impulse are believed to apply uniformly to nerves and muscles, as far as conduction along one cell is concerned. Conduction from cell to cell across a synapse, however, takes place by means of the liberation and diffusion of a transmitter substance. In the case of neuromuscular transmissions, the nerve impulse arriving at the junction causes the liberation of acetylcholine, which diffuses to the muscular part of the junction and there reacts with an unidentified structure or substance called the receptor. As a result of the interactions between the receptor and acetylcholine, the junctional endplate develops a depolarization response called the endplate potential which, if it exceeds a critical threshold value, depolarizes the adjacent areas of the muscle membrane and is then propagated along the whole length of the muscle cell. *See* ACETYLCHOLINE.

The arrival of an impulse at a given part of a muscle cell is the first of a chain of processes which finally elicit the sequence of mechanical and chemical responses that manifest themselves as a contractile event. Knowledge is not available at present as to how the surface stimulus in the membrane affects the interior of the muscle cell, but the fact is that surface excitation instantly causes full activation of the internal bulk, and causes the events described in the following paragraphs.

The various characteristics of these responses depend greatly on the type of muscle, for example, striated voluntary or skeletal, striated cardiac, and smooth or plain muscle. The striated muscle category embraces muscles which, typically, consist of long multinuclear cells, frequently as long as the entire muscle, and often about 0.1 mm in diameter. They consist of longitudinal fibrils, often 1–2 μ thick, embedded in sarcoplasm. The sarcoplasm also carries mitochondria which are numerous when the muscle is of a type that carries on steady activity and high enough respiration to provide energy from moment to moment throughout its action. The fibrils are cross striated by a periodicity of which the sarcomere is the unit. In a sarcomere, the birefringent A band and the less anisotropic and less refractile I band are found. In the center of each I zone, and often considered as the limit between each sarcomere and the next, is the Z membrane which lends a transverse continuity between individual fibrils. Heart muscle, apart from a different architectonic arrangement of the cells, has essentially the same striated structure. Smooth muscle, on the other hand, forms a heterogeneous group, generally consisting of small spindle-shaped cells *without visible fibrillar and cross-striated* differentiations. Its place in the physiological organization of the organism and its associations with the nervous system are altogether different from voluntary muscle. For the biophysical view, in the main the much slower time scale of smooth-muscle responses will be stressed. A major part of the following discussion will apply to striated voluntary muscle.

Contraction and the active state. A single stimulus causes a single event of contractility called a twitch. Its gross manifestation is a generation and decay of tension, or of shortening, as the case may be. The duration of the cycle varies, but may be as short as 10 milliseconds in faster muscles at body temperature (or even less in insect muscles) or as long as several seconds, as in turtle muscles. The often-studied twitch of a frog sartorius muscle lasts up to 1 second at 0°C. Contraction is the "active" event caused by the stimulus. Relaxation

is merely the ceasing of contraction, and there is reason to believe that in the complete absence of external forces, whether tissue elasticity or the muscle's own weight, the contractile structures might stay shortened, because there are no active internal forces that would affect lengthening.

Apart from the grossly visible event, however, many other phenomena occur. Between the stimulus and the first tension development there is a brief latent period of a few thousandths of the twitch duration, during the later part of which a slight diminution of tension, the so-called latency relaxation, occurs. At about the same time also, the muscle undergoes the E. Ernst effect, which is a minute volume diminution of the order of 0.002% of its volume occurring as a result of some process of chemo- or electrostriction. Other minute volume changes occur later on, correlated with phases of activity metabolism. Other early alterations affect the impedance of the muscle, and its turbidity. The latter may change enormously in certain invertebrate muscles, for example, in those in the mantle of the squid, which become quite opaque in contraction. In striated vertebrate muscles, this optical change is much less, and due to the superimposition of light diffraction by the striated structure, not very suitable for theoretical analysis.

When one stimulation is rapidly followed by another or several others, the resulting mechanical responses may be summated. For this to occur, it is necessary that the refractory period of the muscle, during which a newly applied stimulus is ineffective, be brief compared to the time during which fundamental mechanical activity (the active state) is in progress. This is so for skeletal and smooth muscle. Repeated stimulation in skeletal muscle leads to a sustained contraction named a tetanus, in which, usually, considerably greater tension can be produced than in a twitch. In the heart, the refractory period outlasts the active state, and summation and tetanization are therefore not possible.

Although the development of tension or shortening requires time, so that the peak of a twitch is reached not much before the middle of the duration of the cycle, there is proof that fundamentally mechanical activity is at its full level early after the stimulus. This is schematically indicated in Fig. 1, which shows that the active state reaches full intensity almost immediately, and then remains at a plateau of full strength for a definite time before declining. This implies that macroscopically manifest contraction follows the active state with a certain delay, and that it will not reach its possible *maximum* because the active state is by that time declining. In tetanus, on the other hand, the active state persists and full tension will be reached. This circumstance leads to a practical definition: the intensity of the active state is equal to the *tetanic tension*.

The full understanding of the relation between active state and overt contraction requires a con-

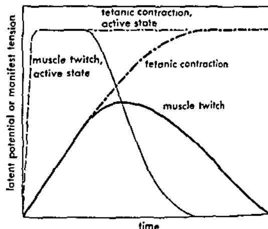


Fig. 1. Schematic diagram of a muscle twitch and of the associated development and decline of the active state, as compared with the development of tetanic contraction and the maintained active state associated with it.

sideration of the mechanics of muscle along the following lines. Muscle is considered to be composed of a contractile component, a series elastic- [illegible] he parallel [illegible] not exclusively, to connective tissue and to the sarcolemmar sheath around each muscle cell. The contractile and series elastic components have not with certainty been identified with morphological structures, and may in part reside within the same cytological elements. Only the contractile component is physiologically activated, and it can exert its tension only through the mediation of the series elasticity. For this to occur, these elastic elements have to be extended until they exert this tension, and this must be done by shortening of the contractile elements. However, this is not instantaneous, but proceeds according to the force-velocity characteristics of the muscle. Hence, the slow ascent of measurable tension in Fig. 1 corresponds to a gradual shortening of the contractile structures, causing an extension of the elastic ones and the generation of tension. At the azimuth of the twitch no shortening or extension occurs, and the measurable tension now corresponds to the momentary intensity of the active state which is already declining.

The length-tension diagram. As in other extensible bodies, muscle can be investigated for the relation between stretch and tension, or between load and stretch. The manner in which the investigation is conducted is dependent upon which parameter is set as the independent variable and which one is measured. Descriptively, the length-tension diagram characterizes a given muscle, but all three components, contractile, series elastic, and parallel elastic, contribute to it. For fundamental biophysical studies, part of the parallel *elasticity* constitutes a problem. The parallel elasticity due to connective tissue could be eliminated, as R. W.

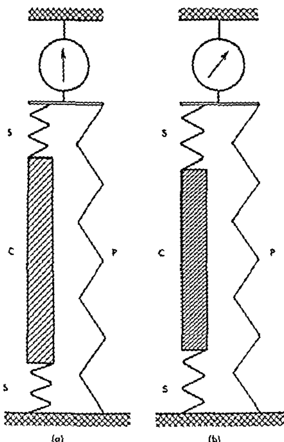


Fig. 2. Schematic diagram of the contractile component (C), the series elastic component (S) and the parallel elastic component (P). (a) In rest, there is no tension and all components are at their rest length. (b) In isometric contraction, tension is developed (as indicated by the measuring gage at the top) because the contractile component shortens enough to stretch the series elastic component so as to transmit the manifest tension.

Ramsey and F. Buchthal did in their important studies, by working with single muscle fibers. This procedure has not generally been followed, partly because of the inconvenience and the possibility of damaging the fiber during isolation, and partly because over wide ranges of shortening, the parallel elasticity has little or no effect. An isolated frog sartorius muscle, for example, exerts practically no tension below its natural length, where many aspects of its contractility can be advantageously studied. At greater length, part of the tension is surely the result of the parallel elasticity.

Figure 3 shows a length-tension diagram of frog sartorius muscle, both in the resting and in the tetanized state. The tetanized state curve shows a more or less pronounced dip at fairly great extension, known as Weber's paradox. It is instructive, as indicated by the dotted curve, to find by subtraction the actively generated tension. This is found to be maximal near the natural length of the muscle; why it falls toward both sides of the natural length is not explained but must be in any future theory of contraction.

The maximal force that can be developed under optimal conditions, the absolute muscle force, differs among the many muscles that have been compared, but not as widely as could have been the case. Most typical skeletal muscles develop a tension of 1–5 kg/cm² of cross section. It is possible that somewhat greater tensions are generated by very specialized muscles. On the other hand, heart muscle develops only about one-tenth of these tensions. It can be calculated, on the basis of Laplace's law, that these tensions suffice to propel the blood against the peripheral arterial resistance.

Estimation of work. The work obtainable from a stimulated muscle can be found from the length-tension diagram as the surface area between it and the abscissa.

$$W = \int_{l_1}^{l_2} P\,dl$$

In this equation W is work, P is tension, l is length, and the integral represents the summation of the length-tension function between two lengths l_1 and l_2. To obtain and measure this work in actual practice is less simple because in order to obtain maximal work, the load must be nearly equal to the tension, meaning that it will barely be raised because upon shortening the tension drops. A. Fick designed variable leverage systems which overcame this difficulty to some extent, and which could now be designed to perfection by means of servosystems. A. V. Hill reported that the muscle can work against inert masses such as a flywheel. However, the most intelligent approach is by means of the Levin-Wyman ergometer, in which a tetanized mus-

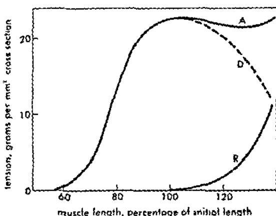


Fig. 3. Length-tension diagrams of muscle, derived from actual measurements on a frog sartorius muscle, expressed in such units as to make them roughly representative for muscle in general. R and A are the length-tension relations for resting and active muscle, respectively. D is the difference between A and R, representing the tension that is actively developed over and above the resting tension.

cle is permitted to shorten at a determined speed over a determined distance, while the tension is being recorded. Because of the equality of action and reaction, the work integral so recorded represents that performed by the muscle at that set speed.

Efficiency. Such measurements of work would assume additional interest by comparison with the total energy output. The term efficiency commonly used for this comparison (also in other instances of biological work) is analogous to the efficiency of a heat engine, $E = W/\Sigma Q$, and would be written as $E = W/\Sigma(\Delta H)$, in which E is the efficiency, W the work performed, ΣQ the total heat evolved, and $\Sigma(\Delta H)$ the sum of the heat effects of the contributing reaction. This use of the term efficiency, unlike that in the case of a heat engine, has no definite theoretical foundation; it means nothing more than a measure of how much work can be obtained at what energetic expenditure. Not much discussion will arise, for over-all respiratory metabolism, as to whether this efficiency ought to be based on the enthalpy or free-energy changes, because these are numerically almost equal. In this sense, the best efficiency a muscle can achieve is about 20–25%. It is known that in certain types of muscle about one-half of the energy is set free during activity as the initial heat and one-half during recovery as the delayed heat. Because delayed heat is associated with the reversal of initial metabolic processes by reactions not directly involved in contraction, the efficiency can be computed or determined for the initial phase separately, and would be about 40–50%. It can be measured by finding the optimal conditions of work performance on the ergometer, and determining the initial heat Q_i experimentally

$$E = W/(W + Q_i)$$

In this equation E is the efficiency and W is the work performed. This was done by Hill on frog sartorius muscle; and 40–45% was the best efficiency found. This is obtainable at a certain optimal speed of contraction, and it is assumed that this speed is a constant characteristic for a given muscle, to which natural motions will tend to adjust themselves for greatest economy. It is possible, on the other hand, that heart muscle can adapt its contraction velocity for optimal efficiency in relation to changes in heart rate. *See* ENTHALPY.

Heat production. The heat measurements just referred to have been perfected by Hill by means of specially designed thermopiles and galvanometers, and later with photoelectrical and electronic amplification. A typical white muscle such as frog sartorius performing an isometric twitch liberates an initial heat of about 0.003 cal/g and a recovery heat of the same order of magnitude. The recovery heat consists of the summated heat effects of those oxidative reactions which reverse the chemical processes that occurred in the initial phase. Anaerobically, the recovery heat is very small, because glycolysis is more nearly thermoneutral than is respiration. The initial heat occurs fundamentally during the contraction phase (preceding the measurable contraction, but coinciding roughly with the active state), showing that the over-all process of activation is exothermic. Its amount represents the total enthalpy change of all the initial reactions, as far as not converted into external work. Experimentally a relaxation heat will occur if any external work is done upon the muscle in relaxation, but fundamentally relaxation is approximately thermoneutral. The initial heat still consists of two parts, one involved in the generation and maintenance of tension, and one which is strictly proportional to any shortening that occurs, regardless of its velocity. The total mobilized energy E, then, consists of maintenance heat A, shortening heat ax proportional to the shortening X, and work W:

$$E = A + ax + W$$

This means that any actual work accomplished involves additional generation of metabolic energy over and above what is used for shortening and activation (W. O. Fenn effect). This important feature seems to have its counterpart in the fact that any work done upon contracting muscle by stretching it does not appear as heat (the B. C. Abbott-X. Aubert effect). The Abbott-Aubert effect must mean that negative work suppresses or reverses biochemical reactions, whereas the Fenn effect indicates mobilization of extra metabolism, of which the $\Sigma\,\Delta H$ equals the work done. The mechanisms of these effects are not understood, but they are so essential that no theory can be considered seriously if it cannot explain them.

In tetanic activity the activation heat is liberated over and again, and is then called maintenance heat. The power exerted for maintaining a tension is inversely defined by its economy. Smooth muscle may be described as having a high economy, so much so that it has often been ascribed to catch mechanisms locking the muscle at a shortened condition without energy expenditure. In general, fast muscles have a low economy, slow muscles a high economy, a relation which is not self-evident on mechanical grounds, but which must result from the limited possibilities available to the working mechanism, and must have a kinetic basis as yet unrecognized.

Some confusion exists in physiological circles owing to the fact that a muscle must expend energy in order to maintain a load, without performing external work. Because physiologically such effort is associated with effects of increased metabolism and fatigue, it is argued that such muscles "work" and that somehow the physical definitions are at fault. This is obviously a gross misunderstanding, as may be illustrated by comparison with inorganic machineries that have to dissipate power without doing work, merely to maintain a load; for example, an electromagnet holding a block of iron, or an

automobile held on a grade with the engine running against a loosely coupled clutch.

Force-velocity relation. When an inert mass such as a flywheel is accelerated by a constant force such as that of a load P falling over a height h, the work performed by the latter is independent of the velocity with which, dependent on its leverage upon the wheel, it is permitted to fall, $Ph = \frac{1}{2}q\omega^2$. By contrast, when muscles are used, either isolated stimulated ones, or for example, the human arm pulling with full effort, the work they can do on an inertia increases as the motion becomes slower. This result must mean that there is some inverse relation between force and velocity of shortening, and was first explicitly formulated as such by W. O. Fenn and S. B. Marsh. It can be argued that this relation reflects some fundamental kinetic property of the contractile matter, related perhaps to the Fenn effect in its mechanism, but likewise entirely uninterpreted. It speaks against the earlier concept of active muscle as a spring with suddenly increased tension, because a spring (apart from friction and acceleration effects) has no force-velocity relation

Several mathematical formulations have been proposed for the relation between force and velocity of shortening. One of these by Hill follows:

$$(P + a)(V + b) = (P_0 + a)b$$

or

$$(P + a)v = (P_0 - P)b$$

Hill's equation has the advantage of easy physical interpretation. Its right part in the second version reveals the significance of the excess force $(P_0 - P)$ in determining v. The left part can be written as $(P + a)\ dx/dt$, in which x is the distance shortened. This reveals that apart from the activation energy the total energy consists of the work Px and an extra energy ax, proportional to the shortening by a factor a. This is in general equal to the shortening heat constant a determined experimentally This connection between dynamic and thermal phenomena, which must have a common basis at the level of molecular kinetics, is an important feature of the Hill theory.

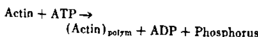
The circumstances under which the force-velocity relation is displayed in the measurable contractions of muscles require review. When muscle shortens under a constant load, this load (assuming acceleration to be negligible) equals the force exerted by the muscle, and the corresponding shortening velocity will directly result. If the muscle was preloaded to this extent, then this shortening will occur directly, apart from the latency period but will be characteristic for the preloaded and therefore stretched muscle. In case of afterloading (that is, the muscle is under no, or slight, initial tension, the weight resting on some support from which it must be lifted), time is first needed before the muscle has developed enough external tension to lift the load. Therefore, a delay time occurs, which becomes longer as the load becomes heavier, and during which the active state is maintained and the contractile component extends the series elastic component until the total tension reaches equality with the load. The purest demonstration of the force-velocity relation is obtained in this fashion, provided that the active state is maintained long enough by tetanization. In isometric contraction, the force-velocity relation determines the rate at which the contractile component can stretch the series elastic component to transmit the tension macroscopically, as was alluded to in the discussion of Fig. 1. One would expect the relation between extension velocity and force, when stretching a stimulated muscle, to appear as a negative continuation of the force-velocity curve. This extension has been treated, among others, in M. J. Polissar's theory, but experimentally the case is difficult because of the possible occurrence of plastic yielding.

Coupling biochemical and mechanical events. In the preceding sections, the biophysical behavior of active muscle has been sketched without special reference to the finer mechanisms responsible for its mechanical alterations Before approaching this problem in detail, it must be stated that energy is required for the activation of muscle and for its shortening and work performance, and that this energy, as always in biological systems, must originate from metabolism. In muscle, this can be aerobic or anaerobic metabolism (respiration or glycolysis, respectively) Biochemistry has shown that both categories of processes seem to be directed exclusively toward the synthesis of energy-rich phosphate compounds, in the form of adenosinetriphosphate (ATP). *See* SPECIALIZED TISSUE. Presumably, then, mechanochemical coupling would consist of an interaction between ATP and some component of the contractile structure, by a process of such specific topochemical nature that contraction results. This process has not been identified, yet experimental progress along these general lines has been rewarding. There are two types of views as to the moment in the contraction cycle at which this energization occurs. The most straightforward explanation is that metabolic energy is used to activate the muscle, and the experimental result that contraction is exothermic from its first beginning is indicative of this. However, a certain popularity has been gained by the opposite type of theory, in which the contraction event is held to be spontaneous and endothermic but happens to be concealed by unknown exothermic factors, with reenergization occurring in relaxation. Although less likely to the experimentalist, such views cannot be definitely excluded. Certain of these theories assume that muscle contracts for thermokinetic reasons, like stretched rubber or heated collagen. This type of theory, if taken literally, is directly disproved by the thermal measurements of Hill and of Aubert which show that contracting muscle, or muscle in rigor caused by iodoacetate poisoning, have normal rather than

rubberlike elasticity. *See* ADENOSINETRIPHOSPHATE (ATP); METABOLISM.

Structure proteins. In the sarcomere, the A band is distinguished by its optical anisotropy, which is of such a nature as to suggest a submicroscopic structure of thin rods embedded in a homogenous medium. This idea has been confirmed in considerable detail by the results of modern electron microscopy. A. Von Muralt and J. T. Edsall identified the protein constituent of this filamentous structure as myosin, on the basis of the fact that solutions of this protein display flow birefringence. When A. V. Engelhardt and M. N. Ljubimova discovered that myosin is active as an ATP-splitting enzyme or ATPase, a direct connection between contractile structure and metabolism suggested itself, although the further development of this problem led to considerable complexity. *See* ENZYME, MICROSCOPE, ELECTRON; PROTEIN.

It was discovered in A. Szent-Györgyi's laboratory that the contractile material consists in the main of two proteins, myosin and actin. According to ultracentrifuge and light-scattering studies of W. Mommaerts, myosin exists in 0.5 *M* potassium chloride solution in the form of anisometric molecules which have a molecular weight of 420,000, may be rod-shaped, and have a length of 1600 angstrom units. Actin can exist in two forms, a monomer (globular or G-actin) and a fibrous form (F-actin). According to F. B Straub, F-actin is formed from the monomer in presence of the salt and ATP. Mommaerts reported that this transformation consists of a linear polymerization of the monomers (molecular weight 60,000) in a stoichiometric reaction with ATP:



$$\text{Actin} + \text{ATP} \rightarrow (\text{Actin})_{\text{polym}} + \text{ADP} + \text{Phosphorus}$$

The stoichiometry of this reaction is such that, should it occur in muscle, it might account for the order of magnitude of the chemical transformations in a single contraction. However, the actual occurrence of these processes in vivo remains to be demonstrated. *See* STOICHIOMETRY.

In solution, F-actin and myosin can combine to a complex, actomyosin, which consists of large molecular swarms (Fig. 4). In actomyosin, in absence of orienting influences, the myosin molecules are attached randomly to the actin strands. Addition of ATP to solutions of actomyosin causes a dissociation of the protein complex into its constituents. In precipitated actomyosin, however, this dissociation cannot occur, and instead ATP causes contraction phenomena. If the actomyosin is merely present in a flocculent form, this contraction reveals itself in a syneresis of the flocs, a process called superprecipitation by Szent-Györgyi. Threads of actomyosin gel contract, especially when well oriented. The problem can also be approached from another aspect and, without disrupting the actomyosin structure, the bulk of other materials can be removed from a muscle fiber by prolonged extraction with aqueous glycerol at low temperature These glycerol-extracted fibers contract vigorously with ATP, developing about the same tension as the original muscle and, as studied by H. H. Weber, display several characteristic features of muscle. Relaxation of such contracted fibers can also be caused by ATP in the presence of a so-called relaxation factor which may be of a complex nature, containing factors assuring the maintenance of ATP, besides a specific relaxation-causing agent.



Fig. 4. Schematic representation of the structure of an actomyosin swarm as occurring in solution. Myosin molecules (rods) are randomly attached to the filament of polymerized actin.

Localization of actin and myosin. The electron miscroscopy of ultrathin muscle sections has added considerable detail to the knowledge of muscle structure. The submicroscopic structure of the sarcomere is shown in diagrammatic form in Fig. 5. This concept may be modified by further work. The sarcomere consists of two systems of interdigitating rods. Heavy rods extend through the full length of the A band, and consist of myosin.

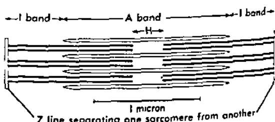


Fig. 5. Schematic representation of the submicroscopic structure of a sarcomere, based upon the electron microscope studies by H. Huxley and J. Hanson. A, Z, I, and H indicate the different microscopically distinguishable regions of a cross striation. Black rods are considered as actin, white rods as myosin.

Because these rods are at least 1 μ long and some 100 or more angstroms thick, they cannot represent single myosin molecules as existing in solution, but must be regular arrays of these, perhaps in the form of crystalline assemblies. Thinner rods, thought to consist of actin, stretch from the Z line throughout the I band into the A band, but not through the full length of this. Contraction is ascribed to a sliding of the actin rods further into the A-band region. This conclusion is based upon phase-contrast and interference microscopy observations. The forces responsible for this sliding are not known; however, further knowledge of the cross bridges reported between the myosin and

actin rods will have a considerable bearing upon this problem. Studies with fluorescent antibodies for myosin and actin have grossly confirmed this localization of the two proteins, but there are significant discrepancies in point of detail, which need further study.

Theory of muscle contraction. Numerous theories have been proposed at one time or another to explain the mechanism of muscular contraction, but invariably these have not taken sufficient account of some or all of the experimental data to be regarded as final. Major factual problems still unsettled concern whether contraction primarily involves a breakdown of ATP or whether other factors intervene; whether contraction consists exclusively of a sliding mechanism or whether the molecular transformation between G- and F-actin occurs in contraction; and whether any role is played by the relaxation factor. The final theory will have to account for such properties as the shortening heat, the Fenn and the Abbott-Aubert effects, and the force-velocity relation. However, theories as considered by F. E. Buchthal, M. Polissar, H. Huxley, M. F. Morales and others constitute a significant step in the final establishment of this knowledge. Of equal importance will be the further study of the link between excitation and the release of contractile and metabolic activity. *See* MUSCLE; MUSCULAR SYSTEM. [W.F.H.M.M.]

Bibliography: X. Aubert, *Le couplage énergétique de la contraction musculaire*, 1956; I. Banga et al, *Myosin and Muscular Contraction*, Stud. Inst. Med Chem., Szeged, vol. 1, 1942; F Buchthal, E. Kaiser, and P. Rosenfalck, *Biol. Medd. Dansk. Vid Selsk*, 21:7, 1951; European Regional Conference on Electron Microscopy, *Proceedings of the Stockholm Conference*, 1956; W. O. Fenn and S. B. Marsh, Muscular force at different speeds of shortening, *J. Physiol.*, 85:277–297, 1935; A. V. Hill, The effect of nitrate, iodide, and bromide on the duration of the active state in skeletal muscle. *Proc. Roy. Soc.* B, 143:81–102, 1954; A. V. Hill, The heat of shortening and the dynamic constants of muscles, *Proc. Roy. Soc.* B, 126:136–195, 1938; H. E. Huxley, The double array of filaments in cross-striated muscle, *J. Biophys. Biochem. Cytol.*, 3(5):631–647, 1957; W F. H. M. Mommaerts, Investigation of presumed breakdown of adenosine-triphosphate and phosphocreatine during a single muscle twitch, *Am. J. Physiol.*, 182: 585–593, 1955; W. F. H. M. Mommaerts, *Muscular Contraction, A Topic in Molecular Physiology*, 1950; M. F. Morales et al., Elementary processes in muscle action: examination of current concepts, *Physiol. Rev.*, 35:475–505, 1955; D. W. Richards and A. P. Fishman (eds.), *Basic Concepts in Cardiovascular Function*, 1959; Society for Experimental Biology, *Fibrous Proteins and their Biological Significance*, Symposia. no. 9, 1955; A. Szent-Györgyi, *Chemical Physiology of Contraction in Body and Heart Muscle*, 1953; A Szent-Györgyi, *Chemistry of Muscular Contraction*, 2d ed., 1951; H. H. Weber and H. Portzehl, Kontraktion, ATP-Cyclus und fibrilläire Proteine des Muskels, *Ergeb. Physiol. biol. Chem. u. exp. Pharmakol.*, 47:369–468, 1952.

Muscovite

One of the mica group of minerals, also called white, or potash, mica. The composition is $K_2Al_4(Si_6Al_2)O_{20}(OH)_4$.

Muscovite occurs in some granites and is especially abundant in pegmatites, the chief commercial deposits. It is a widespread constituent of slates, phyllites, schists, and gneisses. Secondary fine-grained muscovite (sericite), usually of hydrothermal origin, replaces many silicates, particularly feldspars. Detrital muscovite persists in sandstones.

Minor amounts of sodium (Na), barium (Ba), and rubidium (Rb) may substitute for potassium (K); some magnesium (Mg), ferrous iron [Fe(II)], ferric iron [Fe(III)], manganese (Mn), and chromium (Cr, in fuchsite) for aluminum (Al); and minor amounts of fluorine (F) for hydroxyl (OH).

Most muscovites have the two-layer monoclinic structure, a few have the one-layer monoclinic type or the three-layer hexagonal type. Crystal

orientation; plumose to vermicular intergrowths with quartz; globular to hemispherical clusters of concentric plates; and fine-grained masses of irregular or pseudomorphous form. Physical properties include easy perfect basal cleavage to sheets that are flexible and elastic, specific gravity of 2.7–3.1; hardness of 2–2.5. Colors in sheets are in shades of brown, green, and yellow; viewed microscopically muscovite is colorless. *See* MICA; SILICATE MINERALS. [E.W.H.]

Muscular dystrophy

A group of diseases characterized by degeneration or injury to individual muscle cells and not primarily involving the nerve supply. Lesions are found distributed in random fashion in a muscle so that it may contain living, injured, dying, and dead cells, together with some cells which have enlarged in compensatory hypertrophy. Various histologic patterns of injury occur and may be accompanied by an inflammatory reaction to the developing necrosis, or death with dissolution. Grossly, the muscles appear shrunken, soft, flabby, and pale; replacement of the dead muscle tissue by scar tissue sometimes gives it a streaked appearance.

The exact causes are so far unknown for this group of primary hereditable degenerations. A number of distinct clinical types are recognized by the genetic pattern displayed, the muscles involved, and the predominance of shrinkage (atrophy) or enlargement (hypertrophy) which appears early in the disease. *See* HUMAN GENETICS.

Generalized familial muscular dystrophy (Duchenne's dystrophy) is a severe form with early hypertrophy, then atrophy, of muscle groups, which occurs in males under 10 years of age. It runs a progressive downhill course and terminates in early death. Mild restricted muscular dystrophy affects males and females equally and is a slowly progressive disease of both infants and adults. Dystrophica myotonia is a progressive involvement, usually of small muscles of the hand, lower arm, and leg, that typically begins in the third decade of life and runs an extended course. Progressive muscular dystrophy may be confined to certain muscle groups or may be generalized. In all of these diseases, an inborn error of metabolism has been postulated since there are often characteristic biochemical changes, notably in the creatine-creatinine cycle. *See* CREATINE-CREATININE.

It must be understood that atrophy (a form of dystrophy) may also occur as a result of central nervous system disease, as in a stroke, or following damage to the spinal cord, as in poliomyelitis. Other local and systemic disorders may also produce atrophy but these are not true muscular dystrophies. Finally, damage to any nerve innervating a muscle will result in paralysis or atrophy of that muscle; such disorders are of neurogenic origin. The muscular dystrophies are of myogenic origin since the defects occur in the muscle itself, without reference to other related structures. *See* MUSCLE. [E.G.ST.]

Muscular system

The muscular system consists of muscle cells, the contractile elements with the specialized property of exerting tension during contraction, and associated connective tissues. The three morphologic types of muscles are voluntary muscle, involuntary muscle, and cardiac muscle. The voluntary, striated, or skeletal muscles are involved with general posture and movements of the head, body, and limbs. The involuntary, nonstriated, or smooth muscles are the muscles of the walls of hollow organs of the digestive, circulatory, respiratory, and reproductive systems, and other visceral structures. Cardiac muscle is the intrinsic muscle tissue of the heart. In this article the comparative embryology of the voluntary and involuntary muscles of the vertebrates will be outlined, followed by the comparative anatomy of the muscular system.

EMBRYOLOGY

Muscle derivation. The muscles are derived from mesoderm, the middle germ layer. The two exceptions are the dilator muscles of the iris and the myoepithelial cells of the sweat-gland ducts, which are derived from ectoderm. The embryonic mesoderm that differentiates into muscle tissues includes the dorsal mesoderm, head mesenchyme, intermediate mesoderm, and lateral mesoderm. The dorsal mesoderm that condenses into bilateral columns adjacent to the neural tube forms the segmentally arranged myotomes. The intrinsic voluntary muscles of the neck and trunk are differentiated from these myotomes. Some voluntary head musculature (muscles of eye and tongue) and the limb musculature are derived from myotomes in the lower vertebrates but this embryological origin is not readily demonstrable in the higher vertebrates. The voluntary muscles of the branchial (visceral) arches of the head and neck are derived directly from head mesenchyme. The involuntary musculature is differentiated from the intermediate mesoderm (mesomere, urogenital mesoderm, or nephrotomic mesoderm) and the lateral mesoderm (splanchnic mesoderm and hypomere). The intermediate mesoderm differentiates into much of the urogenital system whereas the lateral mesoderm differentiates into the vascular, digestive, and respiratory systems and related structures. *See* EMBRYONIC DIFFERENTIATION; EMBRYONIC INDUCTION.

Differentiation of striated muscles. Striated muscles differentiate from the myotomes of the somites and from mesenchyme of nonmyotomic origin. Muscle cells, fascial cells, tendon cells, and aponeurotic cells are derived from these structures. The premuscle cells (myoblasts) migrate as the organism develops. When the myotome is adjacent to the neural tube, it receives its initial innervation which is retained during subsequent migration of the developing muscles. The innervation of the primordial muscle masses and the retention of this innervation during development is significant as a means of determining the homology of muscles of different species. The nerves are probably not concerned with organizing these muscle masses, because muscles will develop without any innervation in certain monsters. *See* NERVOUS SYSTEM; TERATOGENESIS.

The segmental pattern of the myotomes may be retained in the adult, as in the intercostal muscles of the thorax of mammals and the trunk muscles of fishes. The pattern may be modified as in the flat muscles of the mammalian abdominal wall. The segmental derivation of the muscle masses may be masked by migration as in the eye and tongue muscles, by fusion of muscles from several segments as in the rectus abdominis muscle (the product of the fusion of muscle masses from successive myotomes), by the splitting of muscles into layers as in the flat abdominal muscles, and by fusion and realignment of muscles as in the back musculature of higher vertebrates where muscle fascicles extend through as many as six segments. During metamorphosis in frogs, the segmental patterns of many muscles in the tadpole are altered by the migration, fusion, and splitting of muscle masses to form many muscles in the adult frog.

Axial musculature. In all vertebrates, the muscles of the neck, trunk, and tail are derived from myotomes, although claims have been made that some myoblasts differentiate in dermatomes. The myotomes develop and migrate laterally and ven-

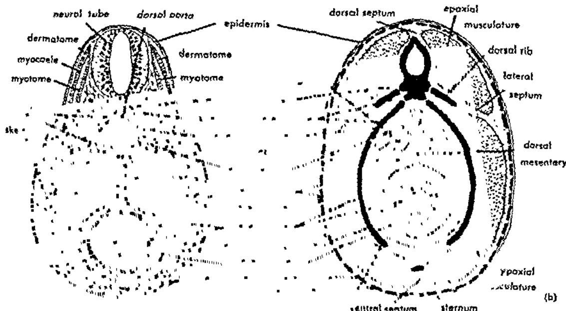


Fig. 1. (,, n) an embryo that give · · ..ues and other structures. Note the location of myotome (dorsal mesoderm), intermediate mesoderm in the region labeled nephrotome, and lateral mesoderm in the tissues surrounding gut and aorta and in the body wall. Schematized cross section of embryo. (b) Diagram illustrating fate of embryonic mesoderm. The myotome differentiates into the epaxial (back) musculature and hypaxial (lateral and ventral body wall) musculature. (From O. E. Nelsen, *Comparative Embryology of the Vertebrates*, Blakiston, 1953)

trally to form the dorsal or epaxial muscle mass (dorsal and lateral back muscles) and the ventral or hypaxial muscle mass (lateral and ventral muscles). The subsequent fate of this migration varies in different animals. In the aquatic animals, whose primary means of locomotion is swimming, the musculature is oriented to produce undulatory motion. Each myotome differentiates into a muscle mass which forms a band from the back to the belly. The muscle cells within each band are directed cephalocaudally. The epaxial musculature (muscle mass innervated by the dorsal ramus of a spinal nerve and located dorsal to the vertebral column) is approximately equal in size to the hypaxial musculature (muscle mass innervated by ventral ramus of a spinal nerve and located ventral to the vertebral column). In most terrestrial animals the bilateral appendages assume primacy in locomotion. In these animals, the myotomes differentiate into the body-wall musculature pattern in which the epaxial musculature is restricted to the back and the hypaxial musculature is present on the lateral and ventral aspect of the body wall. The musculature of the appendages may be of myotomic origin but this is difficult to observe. In animals adapted for aerial locomotion, the primitive segmental pattern is modified. The epaxial musculature is greatly reduced in the trunk but well developed in the nuchal region. The hypaxial musculature is mainly concentrated in the pectoral musculature.

Fate of myotomes. The fate of the myotomes during their differentiation into the axial muscles differs in the various vertebrates. In the fishes the embryonic pattern is retained because the epaxial and hypaxial muscles are relatively equal in size. In the amphibians, two adult patterns are developed. In aquatic amphibians, such as *Necturus*, the embryonic segmental pattern is retained. In terrestial amphibians, such as the frog, the myotomic derivatives are modified during metamorphosis by migration, fusion, and splitting. As a result, the embryonic pattern is modified as a functional adaptation to life on land. In the other terrestrial animals, reptiles and mammals, the embryonic myotomic segmentation is altered further than in the terrestrial amphibians. The epaxial muscles form the erector spinae back muscles. Some hypaxial muscles retain vestiges of the segmental patterns (intercostal muscles of thorax and the rectus abdominis muscle of abdomen), whereas other muscles are the products of migration, fusion, and splitting during development (flat abdominal muscles). The muscle patterns are altered more drastically in birds.

The tail bud mesoderm differentiates into myotomes in tailed animals that exhibit lateral movements. In these forms—fish, tailed amphibians, crocodiles, and whales—the myotomes differentiate into well-developed epaxial muscles and hypaxial muscles that retain their embryonic metamerism. The musculature of animals with prehensile tails

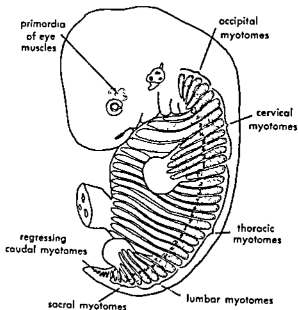


Fig. 2. Diagram of human embryo to illustrate the regions into which myotomes extend. Each stippled area indicates the relative size of each myotome and the adjacent unstippled area outlines the territory into which each myotome migrates. (*From B. M. Patten, Human Embryology, Blakiston, 1953*)

and tails adapted for grasping and wagging movements are partially the result of the migration of myotomes from the hindlimb area. Of significance in mammals is the derivation of the diaphragm from portions of the neck myotomes and their subsequent migration through the neck and thorax to the thoracoabdominal boundary.

Axial muscles. Each axial muscle in the vertebrates is derived from one or more specific myotomes which are numbered according to the vertebral level (cervical, thoracic, lumbar, sacral, caudal, or coccygeal regions) at which the myotome first differentiated. Because the innervation of each myotome occurs during early development, the number in a region of each spinal nerve is identical with that associated with each myotome. In the different classes of animals, the regional number of a myotome and the spinal nerve varies because the number of vertebrae in any region is not similar in the vertebrate classes. Because of this variability, only the segmental numbers of the myotomes in relation to the adult muscles in man, as an example, will be presented.

Epaxial muscles. The epaxial muscles, the intrinsic extensor muscles of the back, are derived from 29 myotomes—the first cervical segment through the fourth sacral segments inclusive. These include 8 cervical, 12 thoracic, 5 lumbar, and 4 sacral segments. The hypaxial cervical musculature is derived from the 8 cervical myotomes. The myotomes differentiate into a prevertebral portion (immediately in front of the vertebral column), a lateral sheet, and a ventral or rectus column. As a result the prevertebral portion forms the prevertebral muscles; the lateral sheet forms the scalene muscles; and the rectus column forms the geniohyoid and infrahyoid muscles. The diaphragm is derived from the hypaxial division of the third through fifth cervical myotomes from whence it migrates.

Hypaxial muscles. The muscles of the ventral and lateral thoracolumbar wall are derived from the hypaxial divisions of the first thoracic through the first lumbar myotomes inclusive. These myotomes differentiate into a main lateral sheet and a rectus column (ventral edge). The intercostal muscles, the oblique abdominal muscles, and the transversus abdominis muscle develop from the main lateral sheet, whereas the rectus abdominis muscle develops from the rectus column. The quadratus lumborum muscle of the posterior abdominal region is derived from the hypaxial divisions of the first through fifth lumbar myotomes. Although direct evidence is difficult to observe, it is probable that the muscles of the pelvic diaphragm (the muscular floor of pelvis including the coccygeus and levator ani muscles) are derived from the hypaxial divisions of the last four sacral and all coccygeal myotomes.

Paired appendages. Theoretically the paired appendages were derived originally either from gill arches (gill-arch theory) or from fin folds (lateral-fold theory). On the basis of their embryological development in the fishes, the fin-fold theory is favored because in these forms the fin musculature is differentiated from myotomes. The myotomes in the vicinity of the developing fin fold migrate to form the dorsal and ventral premuscle mass in each paired fin of the elasmobranch and teleost fishes. The former muscle mass differentiates into the dorsal, elevator, and extensor muscles and the latter muscle mass differentiates into the ventral, depressor, and adductor muscles. They are innervated by the spinal nerves.

Although the myotomic origin of limb musculature is difficult to demonstrate in the higher vertebrates, it is probable that these muscles are differentiated from limb-bud mesoderm derived from myotomes. One line of evidence to support this concept is that the limbs are innervated by the segmental spinal nerves which in lower vertebrates are associated with myotomes. Thus during evolution the original metameric innervation is retained whereas the myotomic source of the limb muscles is masked.

Most of the limb muscles are derived from the limb-bud mesoderm but some proximal limb muscles are not. Some mesenchyme, not derived from limb-bud mesoderm, migrates toward the limb and differentiates into some extrinsic limb muscles with their origins on the axial skeleton and their insertions on the limb skeleton. The trapezius, rhomboid, levator scapulae, and serratus anterior muscles are in this category. Some of the premuscle masses of the limb-bud mesoderm migrate from the limb

bud toward the axial skeleton to form some of the extrinsic muscles with their origins on the axial skeleton and their insertions on the limb skeleton. In this category are such muscles as the pectoral muscles of the forelimb and the gluteal muscles of the hindlimb. Other premuscle masses of the limb-bud mesoderm migrate to form the preaxial group of muscles (those on the cephalic aspect of each limb) and the postaxial group of muscles (those on the caudal aspect of each limb). Both the preaxial and postaxial muscles are subdivided into a flexor or ventral group of muscles and an extensor or dorsal group of muscles. The innervation of the limb follows similar patterns. For example, in man the upper extremity is innervated by the lowest four cervical nerves and the first thoracic nerve (brachial plexus) and the lower extremity by the lowest four lumbar nerves and the first three sacral nerves (lumbosacral plexus). These spinal nerves grow into the limb buds at an early stage in a cephalocaudal pattern. The preaxial muscles of each extremity are innervated by the more cephalic nerves of its plexus and the postaxial muscles by the more caudal nerves of its plexus. Each spinal nerve divides into a ventral (anterior) branch and a dorsal (posterior) branch. The ventral branches innervate the flexor or ventral group of muscles and the dorsal branches the extensor or dorsal group of muscles.

Head and visceral arch musculature. The muscles of the head and visceral arches are derived from myotomic mesoderm and from nonmyotomic mesoderm. The extraocular muscles of the eye and the tongue muscles are differentiated from myotomes or from a mesoderm that phylogenetically was originally derived from myotomes. The visceral arch musculature in all vertebrates is derived from mesenchyme of nonmyotomic origin.

The extraocular muscles of sharks are derived from the preotic somites of the premandibular myotomes (third cranial nerve), the mandibular myotome (fourth cranial nerve), and the hyoid myotome (sixth cranial nerve). The extrinsic muscles of the eye include six which are found in all vertebrates: the superior rectus, internal (anterior) rectus, inferior rectus, and inferior oblique muscles, which are innervated by the third cranial nerve; the superior oblique muscle, which is innervated by the fourth cranial nerve, and the external (posterior or lateral) rectus muscle, which is innervated by the sixth cranial nerve. In addition, the retractor oculi of many mammals and the quadratus muscle and pyramidalis muscle of birds are in this category. The tongue musculature in the sharks develops from six postotic myotomes that migrate ventrally to the hypobranchial region. In higher vertebrates three postotic (occipital) myotomes appear to provide the mesodermal source of this musculature which is innervated in all vertebrates by the hypoglossal nerve. Because direct myotomic origin of the extraocular muscles and the tongue muscles is difficult to demonstrate in the higher vertebrates, there is disagreement in the literature.

The mesoderm of the branchial (gill) arches is derived from head mesoderm which develops in situ and not from any myotome. The first branchial (mandibular) arch mesoderm differentiates into the muscles of mastication that are innervated by the fifth (trigeminal) cranial nerve. The muscles derived from this mesoderm in the fishes are the mandibular adductor muscle and the first ventral constrictor muscle; in the amphibians, the temporal, masseter, pterygoid, and mylohyoid muscles; in birds the pterygotemporal, temporal, and digastric muscles; and in mammals the muscles of mastication (temporal, masseter, and pterygoid muscles), mylohyoid muscle, anterior belly of the digastric muscle, tensor palatini muscle, and the tensor tympani muscle. The second branchial (hyoid) arch mesoderm differentiates into those muscles innervated by the seventh (facial) cranial nerve. The muscles derived from this mesoderm in fishes are the hyoid gill arch muscles; in amphibians, the subhyoid and mandibular depressor muscles; in birds, the sphincters of the neck and the mandibular depressor muscles; and in mammals the muscles of facial expression and other muscles such as the stylohyoid muscle, stapedius muscle, and the anterior belly of the digastric muscle. The mesodermal derivatives of this arch in mammals migrate to the scalp (occipitofrontalis muscle), ear region (auricular muscle), neck (platysma), and the face (orbicularis oculi, orbicularis oris, and others), *collectively called the muscles of facial expression*. The third visceral (first branchial arch) arch mesoderm differentiates into those muscles innervated by the ninth (glossopharyngeal) cranial nerve. The muscles derived from this mesoderm in fishes are the gill constrictor muscles of this arch and in the higher vertebrates (mammals) the stylopharyngeus muscle and the upper constrictors of the pharynx. The mesoderm of the last three visceral arches (second, third, and fourth branchial arches) differentiates into the muscles innervated by the tenth (vagus) cranial nerve. These are the gill constrictor muscles in the fishes and the lower pharyngeal constrictor and laryngeal muscles in the higher vertebrates.

The sternocleidomastoid and trapezius muscles of mammals are innervated by the spinal cord division of the eleventh (spinal accessory) cranial nerve. These muscles are derived either from the mesoderm of the last visceral arch or from postotic myotomes.

Voluntary skin muscles. The skin muscles, the voluntary muscles that move the skin, are divided into two groups, the muscles of facial expression which are innervated by the seventh cranial nerve, and the panniculus carnosus of the body-wall skin that are innervated by the anterior (ventral) thoracic nerves. The panniculus carnosus is derived from myotomic mesoderm that normally forms the pectoral muscles. This muscle, found in such ani-

mals as the porcupine, dog, cat, horse, and guinea pig, may have both its origin and insertion in the skin of some species or its origin on the greater tuberosity of the humerus and its insertion in the fascia of the skin of the back and thigh of other species.

Involuntary muscles. Involuntary muscles arise independently of the segmental myotomes and visceral arches. The visceral mesoderm differentiates into the mesenchyme that forms the smooth muscles of the digestive system, respiratory system, and many blood vessels. The heart also differenti-

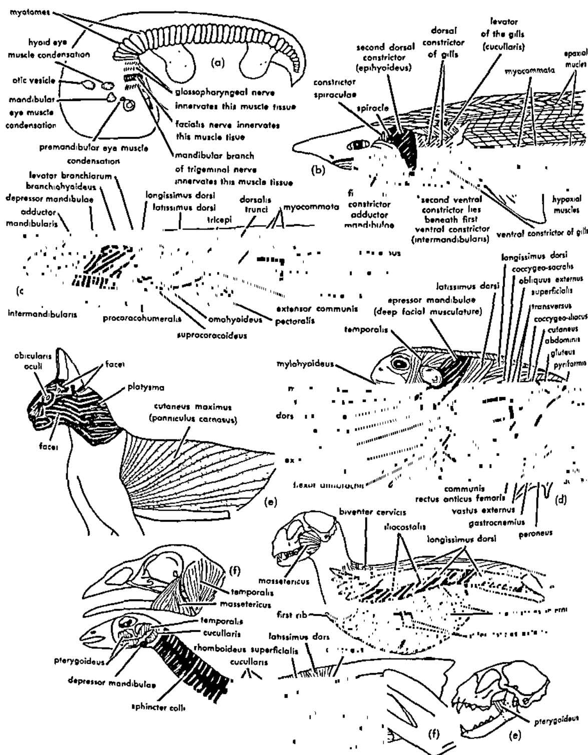


Fig. 3. Development of muscles from the visceral arches and from the myotomes in various vertebrates. (a) Basic areas of the embryo from which voluntary muscles develop. (b) Shark. (c) *Necturus.* (d) Frog. (e) Cat. (f) Goose. (*From O. E. Nelsen, Comparative Embryology of the Vertebrates, Blakiston, 1953*)

ates from this mesoderm. The mesenchyme of the somatopleuric mesoderm of the body wall, head, and limb buds differentiates into the smooth muscles of the blood vessels of these regions. Mesenchyme, whatever its origin, is a potential source of smooth muscle. The smooth muscles of many organs develop in situ from mesenchymal cells. The smooth muscles of the structures of the urinary system and the genital systems are derived from mesenchymal cells of the intermediate mesoderm (nephrotomic mesoderm). The smooth muscles of ectodermal origin are the dilator muscle of the iris and the myoepithelial cells of the ducts of sweat glands.

HISTOGENESIS OF MUSCLE

Embryonic muscle cells (myoblasts) are derived from mesenchymal cells of mesodermal origin.

Voluntary muscle. The myoblasts of voluntary muscles of either myotomic or nonmyotomic origin are uninucleated spindle-shaped cells with clear cytoplasm. These cells soon elongate, become multinucleated, and possess myofibrils. The first myofibrils are sparse and coarse. Cross striations are formed almost as soon as the myofibrils are visible. The elongated cells have centrally located nuclei and peripherally located myofibrils in the early stages of differentiation. Later when the myofibrils fill the cells the nuclei become peripherally located adjacent to the sarcolemma. In man at the time of birth the muscle cells resemble those of the adult except that the nuclei are rounder, the myofibrils are more slender, and the cross striations are less prominent. The multinuclear striated muscle cells are the result of nuclear mitosis unaccompanied by cytoplasmic divisions. The postnatal growth of muscle cells is by hypertrophy. *See* HYPERTROPHY.

Involuntary muscle. Smooth muscles arise from mesenchymal cells. During embryonic life the cells migrate and concentrate in the vicinity of the epithelial linings of the hollow organs. These myoblasts elongate and orient themselves as in the adult organ Myofibrils are visible in the early stages and become more numerous during later development. [C.R.N.]

HISTOLOGY

Three histological types of muscle are recognized: smooth, striated (skeletal), and cardiac. Smooth and cardiac muscle fibers generally occur in layers in the walls of organs, but striated muscle fibers are usually grouped into distinct entities, the skeletal muscles of gross anatomy. Each muscle is surrounded by a sheet of connective tissue, the epimysium, which is a part of the deep fascia. The epimysium is continuous with the connective tissue that invests the bundles of fibers within the muscle, and even with the connective tissue investing the individual fibers.

In some vertebrates such as the rabbit, red and white skeletal muscles can be distinguished, but in most, an individual muscle contains a variable mixture of red and white fibers. Red fibers contain more sarcoplasm and myoglobin than white fibers, and their contraction is more sustained, less subject to fatigue, and often slower. Red muscles, or muscles containing a preponderance of red fibers, tend to be found in situations where the muscles are particularly active in either moving the body or maintaining posture. Some examples are the diaphragm and other respiratory muscles, and the gluteus maximus. White muscles, or muscles with a preponderance of white fibers, are associated with a more intermittent and often a faster and more powerful movement. The biceps and digital muscles are examples. *See* HISTOLOGY.

MUSCLE MECHANICS AND GROUPS

Muscle mechanics. Muscles are usually arranged so that one muscle or group of muscles will pull a structure in a certain direction, and an opposing muscle or group will pull the structure in the opposite direction. Several sets of terms describe these antagonistic actions. Flexion is the movement of a distal part of an appendage toward a more proximal part; this occurs at the elbow and knee. It also describes the bending of the head or trunk toward the ventral surface. Extension is the opposite movement. Flexion and extension are sometimes also applied to forward and backward movements of the appendage at the shoulder and hip, but because they have been used in conflicting senses by different authors, the terms protraction for a forward movement and retraction for a backward movement are more appropriate. Abduction is the movement of a part away from some point of reference, and adduction is movement toward it. For the appendages, the reference point is the midventral line of the body. Various types of rotary movement occur. For example, rotation of the bones of the forearm so that the palm of the hand faces up is supination, and the opposite movement is pronation. *See* MUSCLE (BIOPHYSICS).

The relationships of the muscles and the loads that they work against to the bones is such that three types of lever systems can be recognized (Fig. 5). In first-order levers, the joint (fulcrum) lies between the load and the muscle pull. The action of the triceps on the forearm when one pushes a table away from himself is an example. In second-order levers, the load and muscle pull are on the same side of the lever, with the load closer to the fulcrum than the muscle. A familiar example is the action of the gastrocnemius in raising the body upon the toes. In third-order levers, the load and pull are again on the same side, but the pull is closer to the fulcrum than the load. One example is the action of the teres major in retracting the appendage. Third-order levers are the most common type in the body. Such a lever is mechanically inefficient in the sense that the muscle must develop a force greater than that of the load being moved, but this arrangement has the advantage of compactness and speed of movement, because the muscle is close to the bone and fulcrum and a small contraction of the muscle can induce a greater movement at the end of the lever.

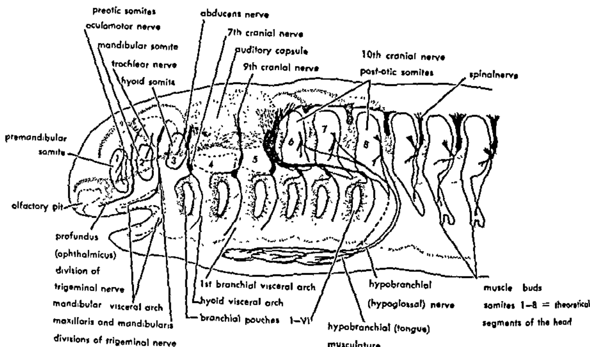


Fig 4. Diagram illustrating the basic plan of the vertebrate head. (From O. E. Nelsen, *Comparative Embryology of the Vertebrates*, Blakiston, 1953)

In general, the power of a muscle is inversely related to the amount and speed of movement that it can cause. Certain patterns of the skeleton and muscles are adapted for extensive, fast movement at the expense of power, whereas others are adapted for power at the expense of speed. In the limb of a horse, which is adapted for long strides and speed, the muscles that move the limb insert close to the fulcrum and the appendage is long. This provides a short power arm but a very long work arm to the lever system (Fig. 5). In the front leg of a mole, which is adapted for powerful digging, the distance from the fulcrum to the insertion of the muscles is relatively greater and the length of the appendage less, with the result that the length of the power arm is increased relative to the length of the work arm.

The inverse relationship of power and the amount and speed of movement is also apparent in the gross shape of muscles and the arrangement of fibers within them. The strength of a muscle is proportional to the number of fibers it contains. The extent of the movement that a muscle can [illegible] which the fibers are arranged in a piniform manner is more powerful than one of equal mass in which the fibers are parallel because it contains more fibers. The longitudinal muscle, however, can induce a more extensive movement because it contains longer fibers. Fan-shaped muscles, because they contain a great number of long fibers, combine to some extent the attributes of piniform and longitudinal muscles.

Muscle groups. Phylogenetically speaking, muscles are very plastic organs. They have undergone considerable change during the evolution of vertebrates, correlated in large part with changes in the organisms' environments and in their methods of support and locomotion. Establishment of homologies among muscles is not easy. Adult relationships can be misleading because muscles have subdivided during their evolution, and parts have migrated far from their original positions. Nerve supply is a more reliable criterion, because nerves have tended to follow the muscles through their evolutionary gymnastics, but often homologies cannot be established without recourse to embryonic development. Determination of the development of the thousands of individual muscles among the vertebrate classes is a monumental task only barely begun. Comparisons of muscles among vertebrates is therefore greatly facilitated if the muscular system is subdivided into groups of muscles whose homology can be more easily established in the various classes.

Muscle groups are particularly distinct in elasmobranchs and other primitive fishes, and they are generally defined on the basis of their embryonic origin in these animals. Two major groups of skeletal muscles are recognized, somatic (parietal) muscles, which develop from the myotomes, and branchiomeric muscles, which develop in the pharyngeal wall from lateral plate mesoderm. The somatic musculature is subdivided into axial muscles, which develop directly from the myotomes and lie along the longitudinal axis of the body, and appendicular muscles, which develop within the

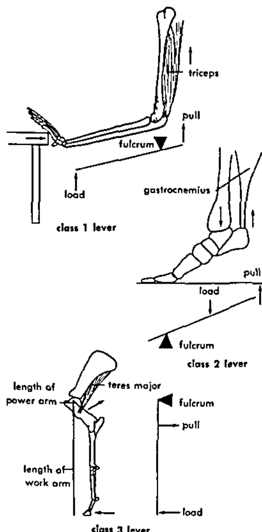


Fig. 5. Types of lever systems found in the human body

limb bud from mesoderm derived phylogenetically as buds from the myotomes.

COMPARATIVE ANATOMY

The vertebrate muscular system is the largest of the organ systems comprising 35–40% of the body weight in man. The movement of vertebrates is accomplished exclusively by muscular action, and muscles play the major role in transporting materials within the body. Muscles also help to tie the bones of the skeleton together and supplement the skeleton in supporting the body against gravity. *See* SKELETAL SYSTEM.

The central part of a skeletal muscle is its belly. The ends of a muscle are attached to the bones by a continuation of its connective tissue into that of the periosteum surrounding the bones. This connective tissue may take the form of a cord-shaped tendon or a broad sheet known as an aponeurosis; or it may be relatively inconspicuous, in which case the muscle is described as having a fleshy attachment. The attachment of the muscle that tends to be stationary as the muscle contracts is its origin; the opposite attachment, which pulls on a structure that can be moved, is its insertion.

Axial musculature. Most of the axial musculature is located along the back and flanks of the body and this part is referred to as trunk musculature. But anteriorly the axial musculature is modified and assigned to other subgroups. Certain of the occipital and neck myotomes form the hypobranchial muscles, and the most anterior myotomes form the extrinsic ocular muscles.

The trunk musculature of cyclostomes consists of a long series of segmental myomeres, each consisting of many longitudinal fibers attaching onto the myosepta (Fig. 6). Each is folded in such a way as to appear approximately zigzag-shaped on the surface. The arrangement in jawed fishes is essentially the same, but the folding of the myomeres

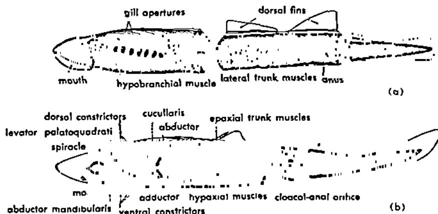


Fig. 6. Superficial muscles. (a) Cyclostome (*Petromyzon*). (b) Elasmobranch (*Squalus*). (*From H. W. Rand, The Chordates, Blakiston, 1950*)

is more complex, and each is divided by a horizontal connective-tissue septum into dorsal (epaxial) and ventral (hypaxial) portions. A spinal nerve passes to each myomere, the dorsal ramus going to the epaxial portion and the ventral ramus to the hypaxial portion. This pattern of innervation persists in all higher vertebrates.

Fishes swim by lateral undulatory movements of the trunk and tail. Contraction of the myomeres is initiated anteriorly, and a given myomere contracts just after the one in front of it. Thus waves of contraction, affecting first one side and then the other, pass down the body. The movement of a given segment of the body is not a simple side-to-side motion, because the inclination of the body changes during the side-to-side motion in a way that gives an effective thrust backward and downward (Fig 7). This is analogous to the thrust given an oar in sculling.

A profound change in the method of locomotion began in ancestral amphibians and has continued through the reptiles to the birds and mammals. Thrusts of the trunk and tail against the air are ineffectual and are replaced by thrusts of the limbs against the ground. The relative mass of the trunk musculature decreases during the evolution of tetrapods and its segmentation becomes obscure, but much of the musculature persists and assumes functions in addition to its original one of bending the trunk.

Epaxial musculature. The epaxial musculature remains powerful in most cases. In amphibians it consists of a group of medial and deep fibers that interlace the vertebrae, and a larger group of superficial fibers (dorsalis trunci). Segmentation is retained and undulations of the trunk and tail still play a role in the locomotion of many amphibians (Fig. 8).

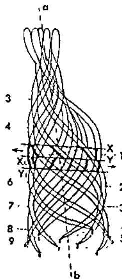


Fig. 7. How an eel swims. Successive drawings of a swimming eel have been superimposed upon the same longitudinal axis (ab). As the body moves from side to side across its longitudinal axis it is inclined (heavy black lines between XY and X_1Y_1) so as to thrust backward. The tip of the tail enscribes a figure 8. *(After Gray from J. Z. Young, The Life of Vertebrates, Oxford University Press, 1952)*

In typical reptiles, the epaxial musculature is more complex. A medial and deep group of small, largely segmental muscles bind the vertebrae together and constitute the transversospinalis system; more laterally the musculature is arranged in two more extensive longitudinal groups, the longissimus dorsi, which lies dorsal to the transverse processes, and the iliocostalis, which is attached to the ribs.

These three main divisions persist in mammals, but posteriorly there is a union of the iliocostalis, longissimus, and sometimes the more superficial part of the transversospinalis system to form a powerful sacrospinalis complex that helps to support the vertebral column. In mammals the body is held off the ground by the limbs; thus the backbone is analogous to a girder supported anteriorly and posteriorly by pillars. Much of the epaxial musculature functions as tie members resisting tension stresses along this girder. Anteriorly there is a cleavage of the epaxial divisions into a host of muscles associated with the complex head and neck movements.

In birds, the epaxial musculature in the trunk is greatly reduced, correlated with a fusion of many of the trunk vertebrae.

Hypaxial musculature. The hypaxial musculature of tetrapods can be subdivided into three groups: (1) a subvertebral (hyposkeletal) group located ventral to the transverse processes and lateral to the centra of the vertebrae, (2) the flank muscles forming the lateral part of the body wall, and (3) the ventral abdominal muscles located on each side of the midventral line.

The subvertebral musculature assists the epaxial muscles in the support and movement of the vertebral column. In mammals, it consists of longitudinal bundles—the longus colli in the neck and the anterior thorax, the quadratus lumborum, and psoas minor more posteriorly.

Most of the flank musculature takes the form of broad, thin sheets of muscle that form much of the body wall and support the viscera. The ancestral, segmental nature of this musculature is retained throughout the trunk in salamanders, but is lost in higher tetrapods except in those parts of the trunk where ribs are well developed (Fig. 8). Three layers can be distinguished in the abdominal region of most tetrapods: a superficial external oblique, whose fibers extend caudally and ventrally; an internal oblique with fibers at right angles to the preceding; and a deep transversus abdominis. This pattern is much the same in the costal region, external intercostals, internal intercostals, and a reduced transversus thoracis being present in mammals. In reptiles the pattern is more complex; the external layer is represented by supracostals, external intercostals, and sometimes a subcutaneous muscle.

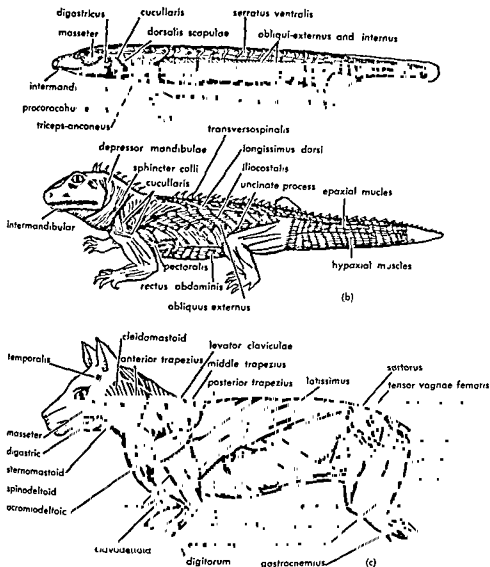


Fig. 8. Superficial muscles. (a) Amphibian (*Necturus*). (b) Reptile (*Sphenodon*). (c) Mammal (*Felis*). (From H. W. Rand, *The Chordates*, Blakiston, 1950)

Respiratory movements of reptiles are accomplished by the costal and abdominal muscles described above, but in mammals, which have a higher metabolic rate, additional respiratory muscles have evolved from the hypaxial muscles: the diaphragm (a derivative of cervical myotomes), serratus dorsalis, scalenes, and transversus costarum. See RESPIRATORY SYSTEM.

Other parts of the hypaxial flank musculature have gained an attachment to the pectoral girdle where they help to transfer body weight from the vertebral column to the girdle and appendage and help to regulate the movement of the girdle. Only a few muscles of this type, the thoracoscapularis and levator scapulae, for example, are present in primitive tetrapods such as salamanders, and the body is not held far off the ground. In mammals, however, this group includes such large and powerful muscles as the serratus ventralis, rhomboideus, and levator scapulae ventralis. In the pelvic region of tetrapods, weight is transferred to the appendage directly across the sacroiliac joint and not by muscles.

The midventral hypaxial musculature in all tetrapods consists of the rectus abdominis, a longitudinal muscle on each side of the midline that extends from the pelvic region to the anterior part of the trunk (Fig. 8). It has evolved from the oblique flank muscles and in some salamanders remains closely associated with them. Transverse tendinous inscriptions are often present and are believed to represent persistent myosepta.

Hypobranchial musculature. The hypo' musculature extends from the pectoral girdle

ward along the ventral surface of the neck and pharynx to the hyoid arch, chin, and into the tongue. It is regarded as a continuation of part of the hypaxial trunk musculature, because it develops ontogenetically in most vertebrates from the ventral portion of several occipital and neck myotomes that grow around the back of the gill region and then into the neck and tongue. The innervation of the hypobranchial muscles in amniotes by the cervical nerves and the hypoglossal nerve, which itself has evolved from certain of the spinal and occipital nerves of fishes and amphibians, further indicates the myotomic origin of this group.

The hypobranchial musculature of cyclostomes (Fig 6) retains its primitive myomeric character, but the myomeres fuse to form longitudinal muscles in higher vertebrates. Traces of myosepta are evident in fishes and some amphibians, but these disappear in amniotes.

In all gnathostomes, the hypobranchial musculature can be divided at the level of the hyoid arch into prehyoid and posthyoid groups (Fig 9). The prehyoid group of elasmobranchs consists of a single pair of muscles, the coracomandibulars, extending caudally from the jaw symphysis to attach to the posthyoid group slightly anterior to the pectoral girdle. The posthyoid group consists of a superficial mass, which can be subdivided into a coracoarcural and coracohyoid, extending between the pectoral girdle and hyoid arch, and a deeper mass, the coracobranchials, extending from the pectoral girdle to the ventral surface of the branchial arches. The coracobranchials act to expand the pharynx and gill pouches; the others help to support the floor of the pharynx and help to move the hyoid arch and open the mouth.

During the evolution of terrestrial vertebrates, loss of most of the coracobranchials occurs, along with the reduction and loss of many parts of the branchial arches, but the rest of the hypobranchial musculature remains and becomes modified in correlation with the evolution of a muscular tongue and the more complex problem of deglutition in a terrestrial environment. The prehyoid group of amphibians consists primarily of a geniohyoid extending from the chin to the hyoid, but a few muscle fibers have separated from it and enter the tongue. These represent the beginning of a complex group of muscles that manipulates the tongue of amniotes: genioglossus, hyoglossus, styloglossus, and probably the intrinsic lingualis. The posthyoid group of primitive terrestrial vertebrates consists primarily of a rectus cervicis extending between the ventral part of the pectoral girdle and hyoid arch, but several slips separate from it to go to other parts of the girdle or to the remnants of the branchial arches. In higher vertebrates, the rectus cervicis has split into several muscles acting upon the larynx and hyoid: sternohyoid, sternothyroid, thyrohyoid, and omohyoid.

Extrinsic ocular muscles. The extrinsic ocular muscles develop from the prootic somites (head cavities). All vertebrates have six in common (Fig. 10). The internal, superior, and inferior recti, together with the inferior oblique, develop from the first somite and are innervated by the oculomotor nerve. The superior oblique and external rectus develop from the second and third somites respectively and are supplied respectively by the trochlear and abducens nerves. All of these muscles insert on the eyeball and move it. In addition, most terrestrial vertebrates, with the exception of birds, have a retractor oculi, a cone-shaped muscle lying deep to the recti, that pulls the eyeball deeper into its socket. The retractor oculi has evolved from the posterior rectus and continues to be innervated by

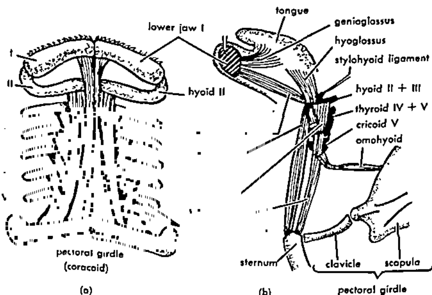


Fig 9. A diagram of the major hypobranchial muscles. (a) Shark, ventral view. (b) Mammal, lateral view. (From H. W. Rand, *The Chordates*, Blakiston, 1950)

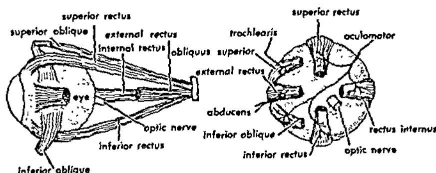


Fig. 10. The human extrinsic ocular muscles. (From H. W. Rand, The Chordates, Blakiston, 1950)

the abducens. In many reptiles and in birds, one or more small muscles have also separated from the posterior rectus, or from the retractor oculi, and act upon the nictitating membrane. A levator palpebrae superioris, present in mammals, completes the ocular group. This muscle elevates the upper eyelid in opposition to the action of certain facial muscles. It has evolved from the superior rectus and is innervated by the oculomotor nerve. *See* EYE.

Appendicular musculature. Limb muscles are often classified as intrinsic, if they lie entirely within the confines of the appendage and girdle, and extrinsic, if they extend from the girdle or appendage to other parts of the body. This scheme has certain merits, but is misleading from the phylogenetic point of view, because the extrinsic muscles are not all appendicular in the sense in which this group has been defined. Some are appendicular muscles that have developed directly from mesoderm in the limb bud, but others are truncal, hypobranchial, and branchiomeric muscles that have secondarily become associated with the girdle. These muscles are considered with the group to which they phylogenetically belong. *See* EVOLUTION, ORGANIC.

Fishes. The paired fins of fishes are primarily horizontal stabilizing keels, but they are also used in deceleration and steering, and to help control

simple. A single dorsal muscle (abductor) and a comparable ventral muscle (adductor) extend from the girdle into the fin (Fig. 6). Certain fibers of these muscles are so arranged that they can protract and retract the fin. The appendicular muscles are supplied by the ventral rami of spinal nerves.

Terrestrial vertebrates. In terrestrial vertebrates, the limbs become the main organs for support and locomotion, and the appendicular muscles become correspondingly powerful and complex. The muscles are too numerous to describe individually, but they can be sorted into dorsal and ventral groups, because tetrapod muscles originate embryonically in piscine fashion from a dorsal and a ventral pre-

*Homologies of appendicular muscles**

Group	*Necturus*	Cat
Pectoral muscles		
Dorsal	Latissimus dorsi	Panniculus carnosus (part); Latissimus dorsi; Teres major
	Subcoracoscapularis	Subscapularis
	Scapular deltoid	Spinodeltoid; Acromiodeltoid
	Procoracohumeralis	Clavodeltoid; Teres minor
	Triceps	Triceps; Epitrochlearis
	Forearm extensors	Forearm extensors
Ventral	Pectoralis	Panniculus carnosus (part); Pectoralis complex
	Supracoracoideus	Supraspinatus; Infraspinatus
	Coracoradialis; Humeroantibrachialis (or brachialis)	Biceps brachii; Brachialis inferior
	Coracobrachialis	Coracobrachialis
	Forearm flexors	Forearm flexors
Pelvic muscles		
Dorsal	Iliotibialis	Sartorius
	Puboischiofemoralis internus	Iliacus; Psoas major; Pectineus; Vasti?
	Ilioextensorius	Rectus femoris
	Iliofibularis	Gluteus maximus
	Iliofemoralis	Tensor fasciae latae?; *Gluteus medius*; *Gluteus minimus*
	Shank *extensors*	Shank extensors
Ventral	Puboischiofemoralis externus (adductor femoris, in *Necturus* not clearly separated from the preceding)	Obturator externus; Quadratus femoris; Adductor brevis et longus
	Pubotibialis	Adductor magnus
	Ischiofemoralis	Obturator internus; *Gemelli*
	Caudofemoralis	*Crurococcygeus* (absent in some mammals); Pyriformis
	Puboischiotibialis	Gracilis
	Ischioflexorius	Semimembranosus; Semitendinosus; Biceps femoris?
	Shank flexors	Shank flexors

* From W. F. Walker, Jr., *Vertebrate Dissection*, Saunders, 1954

muscular mass within the limb bud (see table). In general, the ventral muscles, which also spread onto the anterior surface of the girdle and appendage, act to protract and adduct the limb and to flex its distal segments; the dorsal muscles, which also extend onto the posterior surface of the girdle and appendage, have the opposite effects (retraction, abduction, and extension). The limb muscles also serve as flexible ties or braces that can fix the bones at a joint and support the body.

Amphibians and reptiles. When at rest, the belly of most amphibians and reptiles is on the ground and the proximal segment of each limb extends laterally and slightly dorsally from its articulation with the girdle. Locomotion involves the partial abduction of the humerus and femur to raise the body off the ground as well as their protraction and retraction. Ventral adductor muscles, such as the pectoralis and supracoracoideus in the pectoral region and the puboischiofemoralis externus in the pelvic region, are relatively large and powerful. During the evolution of mammals, the limbs have rotated under the body and extend ventrally from the girdle. The body is held off the ground by bony columns braced by muscles. Adductor muscles are less powerful and certain ones of them have migrated to other positions and have assumed other functions. The supracoracoideus, for example, has extended dorsally onto the scapula to form the mammalian supraspinatus and infraspinatus (Fig. 11). These muscles now act as braces for the limb in its new position and play a role in its protraction and retraction.

Birds. Flight in birds has entailed a considerable modification of the musculature of the pectoral region. As one example, the ventral adductor muscles are exceedingly large and powerful, and the area from which they arise is increased by the enlargement of the sternum and the evolution of a large sternal keel. Not only does a ventral muscle, the pectoralis, play a major role in the downstroke of the humerus, but a ventral muscle, the supracoracoideus, is active in the upstroke as well. The tendon of insertion of the supracoracoideus has shifted so that it passes through a canal between the clavicle, coracoid, and scapula to attach to the upper surface of the humerus. Its action is analogous to pulling down on a rope that passes over a pulley and down onto a weight.

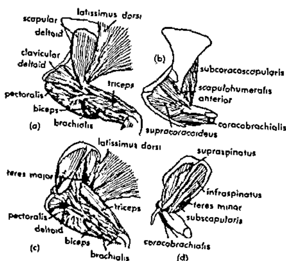


Fig. 11. Lateral views of the shoulder and upper arm muscles. (a, b) Lizard. (c, d) Opossum. Superficial muscles are shown in (a) and (c), deep muscles in (b) and (d). (From A. S. Romer, The Vertebrate Body, 2d ed., Saunders, 1955)

Branchiomeric musculature. The branchiomeric muscles of fishes form a conspicuous part of the muscular system and are rather complex. In jawed fishes they can be subdivided according to the visceral arch with which they are associated. Mandibular muscles act upon the first, or mandibular arch, and are supplied by the trigeminal nerve. The group includes such muscles as the levator palatoquadrati, which in the dogfish helps to support the palatoquadrate cartilage; the adductor mandibulae, the powerful muscles closing the jaws; and the intermandibularis, which together with certain hypobranchial and hyoid muscles, opens the jaws (Fig. 6).

Hyoid muscles act on the second or hyoid arch and are supplied by the facial nerve. The hyoid arch is modified in sharks and many other fishes to help support the palatoquadrate and its musculature is correspondingly modified. The gill slits of bony fishes are covered by an operculum, which has developed from the gill septum of the hyoid arch, and part of the hyoid musculature controls its movement.

The remaining visceral arches (branchial arches) and their muscles are associated with the gills. The muscles of the third arch are supplied by the glossopharyngeal nerve, those of the fourth through seventh arches by the vagus. Muscles of a typical branchial arch include constrictors, interbranchials, adductors, and interarcuals, all of which act to compress the gill pouches and force water out through the gill slits. The gill pouches are opened primarily by the action of the coracobranchials—a part of the hypobranchial musculature. In addition, each branchial arch has a levator, but the levators have united to form a single muscle, the cucullaris, and their insertion has shifted from all but the last branchial arch onto the pectoral girdle (Fig. 6).

During the evolution of terrestrial vertebrates, gills are lost and the visceral arches become reduced and greatly modified. Most of the mandibular muscles remain associated with the jaws and form the various muscles of mastication. In a mammal these are temporalis, masseter, pterygoids, anterior belly of the digastric, and the mylohyoid (Fig. 8). All but the last two close the jaws. The tensori palati, in the soft palate, and the tensor tympani, which attaches to the malleus (a derivative of the mandibular arch), also belong to this

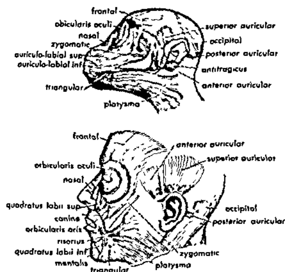


Fig. 12. Facial muscles of two primates, a monkey and man. (From H. W. Rand, The Chordates, Blakiston, 1950)

group. Only a few hyoid muscles remain associated with the hyoid arch or its derivatives: stylohyoid, posterior belly of the digastric, and stapedius. Most of the hyoid musculature has spread out beneath the skin of the face and neck to form the platysma and the numerous facial muscles (Fig. 12). Most of the musculature of the branchial arches is lost, but parts of it form the intrinsic muscles of the larynx and certain pharyngeal muscles. The cucullaris, in contrast, enlarges and subdivides to form the trapezius and sternocleidomastoid, muscles that act on the pectoral girdle and head (Fig. 8). The mammalian motor nerve to these muscles, the spinal accessory, is homologous to part of the vagus of fishes.

Integumentary musculature. In a number of terrestrial vertebrates, particularly amniotes, certain of the more superficial skeletal muscles of the body have spread out beneath the skin and inserted into it. These may be described as integumentary muscles, but it should be emphasized that they are not a natural phylogenetic group but are derived from several different groups.

Integumentary muscles are particularly well developed in mammals and include the facial muscles and platysma, derived from the hyoid musculature, and often a large panniculus carnosus. The last is derived from the pectoralis and latissimus dorsi and fans out beneath the skin of the trunk. The twitching of the skin of an ungulate is caused by this muscle.

Birds and reptiles have a sphincter colli (Fig. 8), a superficial neck muscle derived from the hyoid musculature and hence homologous to the platysma, but they lack facial muscles. Other integumentary muscles, derived from appendicular and trunk muscles, attach to the feathers, especially the large flight feathers on the wings and tails. In snakes, costocutaneous muscles extend from the ribs to the large ventral scales and play an important role in locomotion. [W.F.W.]

Bibliography: F. H. Edgeworth, *The Cranial Muscles of Vertebrates*, 1935; A. S. Romer, *The Vertebrate Body*, 2d ed., 1955; W. F. Walker, Jr., *Vertebrate Dissection*, 1954.

Muscular system disorders

Pathology of the muscular system. Skeletal muscle accounts for approximately 45% of the adult human body weight and is characterized by its striations (transverse bands of alternate light and dark portions) and peripherally placed nuclei. The morphological and functional unit of skeletal muscle is the muscle fiber, which is a large, multinucleated cell, varying in length and width and containing upwards of 60–80 nuclei. Aside from its obvious relation to the skeleton, skeletal or striated muscle is also present in the tongue, the wall of the upper two-thirds of the esophagus, and the diaphragm. Cardiac (heart) muscle is also striated, but has other characteristics which distinguish it [illegible]

a mass of undifferentiated protoplasm, known as sarcoplasm, bound by a cell membrane, known as the sarcolemma, in which myofibrils are embedded. Each muscle fiber contains a large number of myofibrils (threadlike bodies thought by some to represent the contractile elements of muscle), which normally are oriented in a longitudinal direction. Although it has not been definitely established, electron microscopy suggests that the myofibrils may be further differentiated into minute filaments. The precise nature of the cross striations remains unsolved.

Nuclei. The muscle or sarcolemmal nuclei, as they are also known, are elongated and oriented parallel to the myofibrils, and in the majority of muscle fibers are situated at the periphery. Occasional internal nuclei are found, especially in the external muscles of the eye. On cross section, an average of 4–8 nuclei are seen in each fiber. The significance of clumps of nuclei is unclear. Such clumps are usually present only in a few fibers in a given field and may represent nuclei of other structures, such as the motor end-plates. The essential living units of the striated muscle fiber are the sarcolemmal nuclei. If these nuclei are destroyed, regeneration does not take place. In responding to injury, these nuclei, if they survive, may show several signs of activation, for example, proliferation by amitotic division, swelling, appearance of intensely staining nucleoli, and migration among the fibrils (increase in internal nuclei). The means by which injury (disease) excites nuclear activity is unknown.

General pathologies of muscle. Like most other tissue, muscle contains connective tissue elements

and blood vessels. Diseases of muscle, then, need not necessarily manifest themselves by changes in the muscle fiber proper, but by vascular changes, inflammatory exudates or infiltrates, granulomata, presence of foreign bodies or larvae, or changes in the amount and content of the connective tissue.

There are several changes within the fiber proper which can and do occur. Fibers may hypertrophy or atrophy. Cross striations may be lost. The occurrence of ringed fibers (owing to eccentrically placed myofibrils) is an uncommon occurrence, as is the presence of sarcoplasmic masses (protoplasm without myofibrils). Phagocytosis (ingestion by special white blood cells) of muscle fibers also occurs. Subtle changes in morphology may become apparent merely by changes in the staining characteristics of a muscle fiber. *See* ATROPHY; HYPERTROPHY.

As in other tissues of the body, primary and secondary neoplasms are found in muscle, and certain endocrine disorders may be manifest by muscle changes. *See* NEOPLASIA.

Classification. A workable classification of muscle disease divides these entities into the proximal and distal muscle syndromes (proximal and distal referring to position in relation to the axial skeleton). The proximal muscle syndromes include the degenerative forms, such as progressive muscular dystrophy, as well as the inflammatory forms represented by polymyositis. On the other hand, the distal muscle syndromes have classically included the neurogenic muscular atrophies and myotonic dystrophy. It is not rare, however, to encounter a case of spinal atrophy manifested primarily by involvement of the proximal musculature. Similarly, inflammatory and degenerative forms occur which involve the distal muscles.

Because of these varied manifestations of muscle disease, the above clinically derived classification should probably be set aside for a more conventional pathological classification which would include (1) congenital defects; (2) inflammatory disorders; (3) degenerative disorders, both primary (dystrophy) and secondary (atrophy); (4) neoplasms; (5) circulatory and traumatic disorders; and (6) obscure pathology secondary to endocrine or metabolic disorders.

Congenital muscular defects. These may include variations in structure, such as differences in size and position (origin and insertion), as well as congenital absence of muscles. Those muscles showing the greatest frequency of defect are those in association with the shoulder girdle, although almost any site can be affected. Other congenital defects such as club foot, torticollis (wry-neck), and congenital elevation of the scapula are due to shortening and fibrosis of the responsible muscles. Microscopically, the chief finding is replacement of normal muscle by relatively acellular connective tissue.

The term myositis is used to describe changes suggesting an inflammatory reaction, such as the presence of polymorphonuclear leukocytes (white blood cells) or other inflammatory cells, an exudate of plasma, and varying degrees of damage to and reaction by the muscle fibers and associated connective tissue. Two general types of myositis may be distinguished, that in which an injurious agent, such as a bacterium, virus, or parasite, is responsible for the changes, although not always demonstrable; and that in which no causative agent can be discovered, such as in polymyositis or dermatomyositis. These latter entities are believed to be closely related to the so-called collagen diseases

a causative agent, the changes observed within the muscle are generally well defined and sometimes quite characteristic. Examples might include the abscess formation with necrosis (death) of muscle fibers associated with streptococcal or staphylococcal infection. Similarly, in trichinosis the larvae may be easily demonstrated microscopically with a limited focus of muscle necrosis, a tendency to walling off by the connective tissue, and a slight to marked degree of infiltration by white blood cells. *See* DEATH; TRICHINOSIS.

Tuberculous myositis may be secondary to direct extension from a primary focus or by hematogenous

may be demonstrated, but they are not found within muscle fibers. Similar lesions without any tubercle bacilli are seen in sarcoidosis. Syphilis may also produce granulomatous lesions in muscle, and numerous fungi are known to produce this type of reaction. *See* SYPHILIS; TUBERCULARIACEAE; TUBERCULOSIS.

In regard to those myositides, such as polymyositis, in which the etiological factor is unknown the microscopic picture generally is determined by the length of time elapsed since the onset of the affliction. In all forms, the most striking microscopic characteristic is the widespread degeneration of muscle fibers. In the acute and subacute forms varying degrees of white blood cell infiltrates are seen, slight hyperplasia of connective tissue elements is apparent, and there may be proliferation of the smaller blood vessels with a thickening of their walls. In chronic cases or during the late stages of the disease, nearly all the muscle fibers have disappeared and in their place dense fibrous connective tissue is seen with occasional fragments of normal-appearing muscle. The over all picture is quite similar to that seen in the degenerative muscle diseases, such as progressive muscular dystrophy, but can be distinguished from the dystrophies by evidences of attempts at regeneration, for example, budding muscle fibers, a process the lack of which is said to distinguish muscular dystrophy from all other myopathies (muscle diseases).

Degenerative myopathies. Degenerative myopathies are of two types, (1) the dystrophies in which the pathology is limited to the muscle without any associated changes in the nervous system and (2) the atrophies which represent muscle disease secondary to primary spinal or neural (peripheral) disease.

Muscular dystrophies. The muscular dystrophies are a group of muscle diseases of unknown etiology which usually affect the proximal musculature (shoulder and pelvic girdles or both) symmetrically. Clinically, several types are distinguished by age of onset, familial history, and group of muscles involved, such as severe generalized familial muscular dystrophy (progressive with onset in childhood), mild restricted muscular dystrophy, such as the facioscapulohumeral dystrophy of Landouzy and Dejerine, and the progressive ophthalmoplegias which, as their name would indicate, affect primarily the external muscles of the eye; however, the pathology is more or less identical. Another dystrophic myopathy is myotonic dystrophy, which involves distal as well as proximal muscles and is also associated with cataracts, premature baldness, and testicular atrophy. Its most characteristic feature is the phenomenon of myotonia which refers to the prolonged contraction of certain muscles with an associated delay in relaxation.

The most striking microscopic feature of dystrophic muscle is the great variation in size of the muscle fibers. Other prominent features include an increase in the fibrous connective tissue as well as the presence of increasing amounts of fat in the connective tissue. Many fibers are greatly enlarged, and the feeling is that this enlargement later proceeds to atrophy. The sarcolemmal nuclei are also swollen and generally increased in number. Although the power for regeneration is lost in advanced cases of myositis and the atrophies (spinal and neural), dystrophic muscle at no time demonstrates this ability. The absence of regenerative activity in this disease then is its fundamental characteristic. Although the muscle nuclei may proliferate and seem to be active, the essential regenerative process is gone and no new muscle fibrils appear. Therefore, nuclear proliferation, although most commonly associated with signs of regeneration and in the past almost uniformly interpreted as such, does not require this interpretation, for in progressive muscular dystrophy when nuclear proliferation is a prominent finding, the other evidences for regeneration, that is, new sarcolemmal tubes, muscle sprouts, new myofibrils, and so on, never follow.

In addition to the findings described for the muscular dystrophies, myotonic dystrophic muscle shows an increase in internal nuclei as well as the presence of so-called sarcoplasmic masses, regions within a given fiber containing protoplasm without myofibrils.

Atrophies. The spinal and neural atrophies represent secondary myopathies resulting from primary disease of the nervous system. Examples include the progressive spinal atrophy of childhood (Werdnig-Hoffman disease), the adult forms (atrophy of Aran-Duchenne), amyotrophic lateral sclerosis, and progressive bulbar palsy. The etiology of these diseases is unknown. The predominant microscopic feature in muscle is fiber atrophy of varying degrees dependent upon the stage of the disease. Distal muscles are involved early, whereas in the end stages both distal and proximal muscles are usually involved. Microscopically, groups of atrophic muscle fibers may be seen interspersed with groups of normal-appearing fibers. A classical form of spinal atrophy is seen in acute anterior poliomyelitis, the etiological agents being the well-known polio viruses. The pathological changes of the muscles are the same as in the previously described atrophies. Similar microscopic changes are seen in cases of neural atrophies where the primary pathology resides in a peripheral motor nerve.

Other myopathies. Two other myopathies which seem to be related to the degenerative muscle diseases are amyotonia congenita (congenital hypotonia or Oppenheim's disease) and central-core disease of Shy and Magee. Amyotonia congenita is characterized by proximal muscle weakness which gradually improves with time. Microscopically, the muscle has an immature appearance with small fibers and an apparent decrease in sarcolemmal nuclei. Etiology is unknown. Central-core disease is a disease with proximal muscle weakness and a characteristic microscopic appearance in which the central portions of the muscle fibers have staining properties which are different from those of the surrounding, more peripheral part of the fiber. No morphological changes have been described and etiology is unknown.

Neoplasms. Primary neoplasms (tumors) of muscle include the rhabdomyoma, rhabdomyosarcoma, and the granular-cell myoblastoma. The rhabdomyoma is quite rare; nevertheless, several types have been described, all of which are considered benign (noncancerous) A characteristic cell type is the so-called spider cell, a large cell with striated fibrils surrounding the nucleus in a concentric arrangement. Rhabdomyosarcoma is the malignant form of this tumor. Pleomorphism (variation in size and shape) is marked and striated cells are rare. The exact nature of the granular-cell myoblastoma is open to controversy; some authorities doubt that the cells of this tumor are muscular in origin. The predominant cell type is a large, coarsely granular cell without striations. Such tumors most commonly arise in the tongue and are generally considered to be benign. Secondary (metastatic) neoplasms are found frequently in muscle, but are not as common as one might suppose considering the extent of skeletal muscle (nearly one-half the body weight) and its excellent vascularization. See NEOPLASIA.

Trauma. Trauma to muscles, if severe, may cause a shattering of the fibers with subsequent necrosis

(death) and hemorrhage. Occlusions of large arteries to the extremities may result in extensive necrosis of muscle with the development of gangrene.

Chronic diseases. Chronic diseases in which emaciation and cachexia are prominent frequently result in muscular atrophy of varying degrees. The most common endocrine myopathy is chronic thyrotoxic myopathy. Muscle weakness and atrophy, primarily in the muscles of the shoulder and pelvic girdles, are associated with thyrotoxicosis, although the muscle changes may precede clinical evidence of thyroid malfunction. Atrophy here merely refers to decrease in muscle fiber size and does not carry the same connotation as in the spinal atrophies.

Other muscle diseases which do not lend themselves to pathological classification include such disorders as myasthenia gravis and periodic paralysis, well-known clinical entities, but without characteristic microscopic changes

In summary then, muscle disease, although quite varied in its clinical manifestations, is generally less variable pathologically, with relatively few microscopic changes which can be considered diagnostic of any particular clinical entity. [G.B.S.]

Bibliography: R. D. Adams, D. Denny-Brown, and C. M. Pearson, *Diseases of Muscle*, 1954; J. G. Greenfield, G. M. Shy, E. C. Alvord, Jr., and L. Berg, *An Atlas of Muscle Pathology in Neuromuscular Diseases*, 1957.

Mushroom

A fungus belonging to the basidiomycetous order Agaricales, or the fruiting body (basidiocarp) of such a fungus. The generalized life history of a mushroom begins with the germination of a basidiospore which produces a primary mycelium composed of filaments (hyphae) with one nucleus per cell. If two compatible primary mycelia fuse, a secondary mycelium with two nuclei in each cell develops. This underground secondary mycelium, which may be extensive and perennial, forms at its periphery small masses of compacted hyphae which enlarge and differentiate into immature basidiocarps, or mushroom buttons (Fig. 1). These buttons enlarge rapidly, burst through the soil and become mature basidiocarps (Fig. 2). Basidia form on plates of tissue (gills) underneath the cap of the mushroom. Each basidium produces basidiospores (usually four) which are shot off and dispersed by wind. Since mushrooms develop near the margin of the circular mycelium, they sometimes form "fairy rings." See BASIDIOMYCETES; FUNGI.

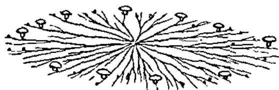


Fig 1. Diagram illustrating relation between basidiocarps and mycelium of a mushroom.

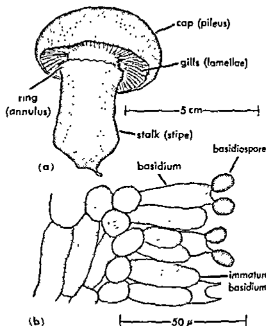


Fig. 2. (a) Basidiocarp of the common cultivated mushroom, *Agaricus bisporus*. (b) Cross section of gill of *A. bisporus*, showing basidia in various stages of development.

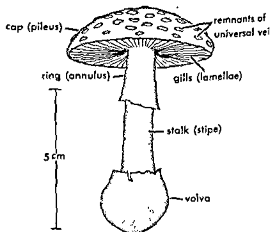


Fig. 3. Basidiocarp of *Amanita muscaria* (poisonous). Poisonous species such as *Amanita* characteristically have a volva.

Basidiocarps of some mushrooms are prized as food; others are poisonous. The term *toadstool* is commonly applied to basidiocarps of poisonous species. The only way to distinguish edible from poisonous species is by accurate identification. Poisoning by such species as *Amanita verna* or *A. phalloides* is fatal in the majority of cases. The poisons (α-amanitine, β-amanitine, phalloidine) are peptides.

Some mushrooms contain hallucinogenic principles. *Amanita muscaria*, which is poisonous when eaten in large amounts, is chewed by natives in central and northern Siberia to produce hallucinations.

and *Psilocybe mexicana* is used in certain parts of Mexico for the same purpose (Fig. 3).

Most mushrooms on the American market are cultivated. Annual production amounts to 60,000,000 lb. Basidiospores of *Agaricus bisporus* are germinated and the mycelium allowed to develop on sterilized grain to form mushroom spawn which is used to inoculate compost. The compost may be horse manure or be derived from straw and leguminous plants which have been chopped, heaped, moistened, turned, and finally placed in shelf beds about 6 in. deep in a cave or building which has provision for control of temperature, humidity, and ventilation. The compost is covered with a 1-in. layer of soil after the mycelium has grown for 3 weeks, and about 1 month later the first mushrooms appear. Successive crops of basidiocarps appear every 4–5 days for 2–4 months.

The mycelium of *Agaricus bisporus* and other

pressed into pellets suitable for flavoring soups and gravies. [R.M.P.]

Bibliography: A. H. Smith, *The Mushroom Hunter's Field Guide*, 1958.

Musical acoustics

That part of acoustics which is relevant to music. The theory of vibrating strings, air columns, and plates, spectrum analysis, frequency measurement, psychoacoustics—even architectural acoustics and electrical circuit theory—all become part of musical acoustics when viewed as potential contributors to the composition, performance, and appreciation of music. The remarks that follow, however, are restricted principally to physical characteristics of sounds that may be heard as music. For information on generators of musical sounds *see* MUSICAL INSTRUMENTS. For a discussion of sound production by the human voice *see* SPEECH.

Beats. A tuning fork, if not driven too vigorously, produces a simple tone (a sine wave); if a second tuning fork sounded simultaneously is of slightly different frequency the result is a sound wave that waxes and wanes in amplitude at a frequency equal to the difference between the frequencies of the two component sounds. These periodic variations in magnitude of sound are called beats, and the rate at which they occur is the beat frequency. *See* BEAT; TUNING FORK.

Consonance. When two musical instruments are sounded together at random, beats may occur—the combination may be rough and dissonant, even though the individual sounds are quite agreeable. It was discovered by the Greeks, long before the Christian era, that when two strings of the same material and tension are sounded together, the result is pleasant if the lengths of the strings are in the ratio 1/1, 1/2, 1/3, or 1/4. Since the frequency of a vibrating string is very nearly inversely proportional to its length, the corresponding statement is made today that intervals between two sounds are consonant if the respective frequencies are in the ratio 1, 2, 3, or 4. H. von Helmholtz explained that with such intervals, consonance exists because beats among the upper partials of the two sounds are minimized. For example, if the frequencies of the partials of the sound of one string are 200, 400, 600, 800, 1000, and 1200 cycles per second (cps), and those of another are 300, 600, 900, and 1200 cps, two of the partials coincide in frequency; that is, the third and sixth partials of the first sound match respectively the second and fourth partials of the other. The ratio of the fundamental frequencies is $200/300 = 2/3$; the interval between the two sounds is called a perfect fifth. If one of the strings were slightly mistuned, beats would occur between the nearly coinciding partials. *See* PARTIAL TONE; VIBRATION.

Intervals of different size differ in their degree of consonance because the amount of beating among higher partials differs. A chord is a combination of two or more tones; whether it is consonant depends upon the frequency ratios of the constituent tones.

A slightly mistuned octave, on most theories of consonance, should be very dissonant indeed; yet, it is certain that, in an orchestra, octaves are not played by all instruments exactly as a 2:1 frequency ratio. Singers typically sing an octave jump as a frequency ratio greater than 2:1. Thus, although existing theories of consonance may correctly indicate trends, they do not explain in detail what is actually accepted as consonance in music. *See* OCTAVE.

Timbre. In discussions of timbre (tone quality), it has long been the custom to state that differences in quality of tone are solely dependent on the occurrence and strength of partial tones. Although H. von Helmholtz, in making this statement, recognized that the characteristic tone of some instruments is dependent upon the way the tone stops and starts, he chose to restrict his attention to the "peculiarities of the musical tone which continues uniformly" and to consider as musical only those tones with harmonic upper partials. Many writers since have adopted these simplifying but not realistic assumptions; according to such simplifications, the piano is not a musical instrument!

The transient parts of a sound contain important clues by which different instruments are identified. A sustained high tone on the clarinet, for example, is practically indistinguishable from the same tone (sustained) played on the flute, but the initiation of the sound is likely to be noticeably different on the two instruments.

Another distinctive characteristic of the tone quality of an instrument is the formant, or frequency range within which the partials of the sounds emitted by the instrument have relatively large amplitudes

Vibrato. Measuring instruments developed in the twentieth century make it amply clear that most of

the sounds employed in the artistic performance of music are not steady. It is true that a relatively steady tone, as from a sustained organ pipe, has a proper place in music. However, the good violinist or singer causes a smooth periodic variation in amplitude and frequency of a tone, or both, at a rate on the order of 6 per second; this variation is called vibrato.

Instruments playing the same note in ensemble do not produce exactly the same frequencies: there is a musically desirable "chorus" effect that has been little recognized in theories of consonance or of musical scales. Even the three unison strings in a piano, struck by a single key, beat together slightly as they decay, and the combination is characterized by a "liveness" that disappears if one listens to a single string alone. Suitable deviation from the regular is an important attribute of artistic performance.

Tuning note A; standard pitch. An important practical matter for musical acoustics is to ensure that instruments of the orchestra are capable of being tuned in unison with each other. This leads to the conclusion that there must be an international standard tuning frequency. The following reasons are given for the adoption of an international standard: (1) the tuning adjustment of wind instruments is more or less limited (even stringed instruments can be harmed by too high string tension); (2) the relative intonation of wind instruments changes as the tuning slide is moved; (3) instruments like the celesta cannot be tuned by the player, and instruments like the pipe organ can be retuned only at considerable expense; (4) musicians play in more than one orchestra, and some may even play in a different state or country each night; (5) international trade is facilitated when a standardized product can be sold anywhere in the world. Even the coloratura soprano should find comfort in knowing that the piano that will accompany her will not be tuned significantly higher than she is accustomed to singing.

By international agreement reached in 1955, the frequency of the note A to which the orchestra is traditionally tuned is 440 cps.

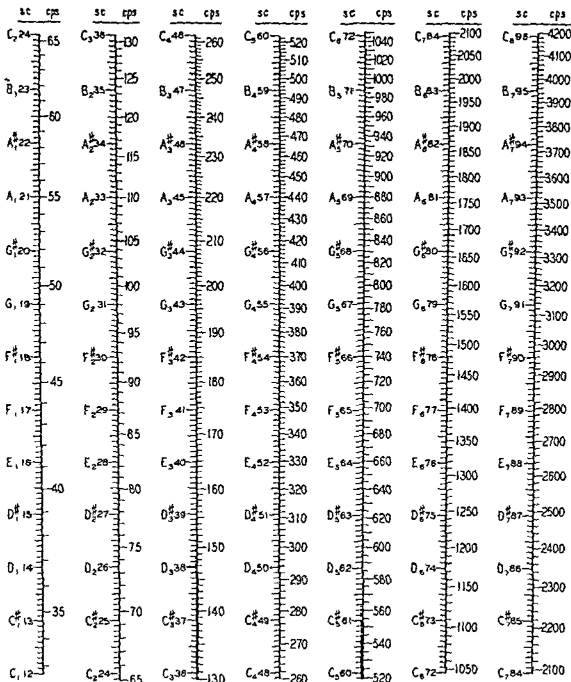
The term "standard pitch" has long been used, but the standard is clearly one of frequency. Pitch is an attribute of sensation which depends on magnitude and waveform, as well as frequency of the stimulus (*see* PITCH). For example, a tuning bar of 440 cps can, if struck vigorously, cause a pitch (perceived by the ear) lower than that caused by a similar bar of frequency 435 cps that is struck lightly; the frequencies, however, are not essentially changed by the manner of striking.

Tuning and temperament. Methods of tuning different musical instruments can lead to somewhat different frequencies for the notes intended to be used, for example, for the diatonic scale. Suppose that the A string of a violin has been tuned to a frequency of 440 cps, and that the E string next to it has been adjusted until beats disappear. As explained previously in connection with consonance, the absence of beats indicates that the interval from E down to A (a perfect fifth) has been tuned so that the frequency ratio is 3/2, and the frequency of the E string is $(3/2) \times 440 = 660$ cps. The frequency of an E an octave lower would be $660/2 = 330$ cps. *See* SCALE (MUSIC).

Pythagorean tuning. An entire scale can be tuned by the use of an ascending sequence of perfect fifths (represented by the frequency ratio 3/2) and by lowering the resulting sounds by a number of octaves (each represented by the frequency ratio 2/1). This is called Pythagorean tuning. The relative frequencies can be found by dealing with ratios only. If the keynote is taken as C, having relative frequency of unity (1/1), the relative frequency of the G above is 3/2. The D a perfect fifth above this G would have a relative frequency of $(3/2) \times (3/2) = 9/4$; a D within an octave of the keynote can be obtained by division by 2; its relative frequency is $(9/4)/2 = 9/8$ (see table). Similarly, the relative frequency for A a perfect fifth above

Major diatonic scale rendered in four different tuning systems (ratios and intervals relate to the first note)

Note	C_4	D_4	E_4	F_4	G_4	A_4	B_4	C_5
Pythagorean								
Frequency, cps	260.7	293.3	330.0	347.7	391.1	440.0	495.0	521.5
Frequency ratio	1/1	9/8	81/64	4/3	3/2	27/16	243/128	2/1
Interval, cents	0	204	408	498	702	906	1110	1200
Just								
Frequency, cps	264	297	330	352	396	440	495	528
Frequency ratio	1/1	9/8	5/4	4/3	3/2	5/3	15/8	2/1
Interval, cents	0	204	386	498	702	884	1088	1200
Meantone								
Frequency, cps	263.2	294.2	329.0	352.0	393.5	440	491.9	526.4
Frequency ratio	1/1	$5^{1/2}/2$	5/4	$2/5^{1/4}$	$5^{1/4}$	$5^{3/4}/2$	$5^{5/4}/4$	2/1
Interval, cents	0	193	386	503	697	890	1083	1200
Equal temperament								
Frequency, cps	261.6	293.7	329.6	349.2	392.0	440.0	493.9	523.3
Frequency ratio	1.0000	1.1225	1.2599	1.3348	1.4983	1.6818	1.8877	2.0000
Interval, cents	0	200	400	500	700	900	1100	1200



SUBSCRIPT OF NOTE NAME INDICATES OCTAVE LEVEL
SEE "TERMINOLOGY FOR LOGARITHMIC FREQUENCY UNITS" J ACOUS SOC AM 11, 134-139 (1939) FOR FORMULAS, REFERENCES.

RWY NCP LEW

Nomogram relating frequency in cycles per second to intervals of the equally tempered scale in semitone count (sc) based on the A of 440 cps, the intervals in semitones being counted from C_0 = 16.35 cps. Note names are identified by a subscript, which is the number of octaves above the octave starting with C_0. (C. G. Conn, Ltd.)

D is $(3/2) \times (9/8) = 27/16$. By use of two more such perfect fifths, all notes of the major diatonic scale are thus accounted for, except F. The way to obtain F is to drop down a fifth from C; this is tantamount to starting the upward series of fifths from F to generate the scale of C. If the note A is assigned the frequency 440 cps, the other frequencies can be found by proportion with the ratios already computed; the result is shown in the table. It can be seen that all the ratios for Pythagorean tuning are ratios of the numbers 2 and 3 raised to different powers, for example. $81/64 = 3^4/2^6$.

Just tuning. Two strings or organ pipes maintained steadily in vibration, if tuned at the Pythagorean interval of a major third in the frequency ratio $81/64 = 1.265$, beat noticeably with each other; this beating can be minimized by changing the ratio to $5/4 = 1.25$. A tuning that includes this interval is called just (one of its various names); it is generated by octave rearrangement of the notes of three consecutive triads, each having the frequency ratio 4:5:6. (Consecutive triads are such that the highest note of one is the lowest note of the next.) The name of the scale is obtained from the middle triad: for the C scale given in the table, the required triads are F-A-C, C-E-G, and G-B-D.

If one tunes a keyboard instrument in just intonation, on playing it will quickly become apparent to the ear that not all thirds and fifths are in the 4:5:6 ratios and thus are not free of beats The supertonic triad D-F-A is quite dissonant because the fifth is too small; this and other difficulties associated with modulation prevent general use of just intonation on keyboard or fretted instruments. On [illegible]

It has long been alleged that a sensitive string quartet or vocal quartet performs in just intonation although not in the rigid sense suggested by the table; the performers are presumed to make slight changes to bring the frequencies in a given chord to the proper ratios. Conceivably the habits of a century ago were different from those of today; experimental evidence is not very complete, but modern attempts at measurement tend rather to suggest intonations nearer the Pythagorean or equally tempered.

Frequency ratios have been included in the table because they give clues to the origins of the intervals. Ratios are not, however, always convenient for comparison; for this purpose, therefore, the corresponding intervals from the first note are also measured in cents. (An octave is an interval of 1200 cents.) In the meantone temperament, for example, it will be noticed that the major third C-E is an interval of 386 cents, being the same as the major third in the just tuning and smaller than the major third of 408 cents in the Pythagorean tuning. In the just tuning, there are two sizes of whole tone, the greater of 204 cents between C and D and the lesser of 182 cents between D and E. The meantone tuning takes its name from the fact that the mean value of $(204 + 182)/2 = 193$ cents is adopted for the whole step interval. The result is a major third relatively free of beats, and a tuning rather pleasing and harmonious; the so-called perfect fifth, however, sounds noticeably flat to present day ears. The meantone tuning can be generated by an upward succession of 6 fifths (each of frequency ratio $5^{1/4}$) starting on F_3, and appropriate lowering by octaves. If the sequence is extended in both directions to generate chromatic notes a "wolf" finally appears as a much too large and dissonant fifth (diminished sixth) of 738 cents that is in contrast with the others of 697 cents, and free modulation is thereby thwarted.

The obvious answer to the demand for freedom of modulation and chromatic notes was to divide the octave into 12 equal parts, and to select from these intervals (see the diatonic scale in the equal temperament section at the bottom of the table). Ways to accomplish this division of the octave had been known for a long time but, after an extended transition period, equal temperament finally replaced meantone temperament nearly a century and a half ago as the preferred tuning for keyboard instruments.

Equally tempered scale. The equally-tempered chromatic scale, based on the A of 440 cps, constitutes a convenient frame of reference in which any frequency can be located (see the accompanying nomogram, from which the approximate frequency of any note can be read). For example, the highest note of the piano, C_8, has a frequency just under 4200 cps. This note is 96 semitones counted upward from $C_0 = 16.35$ cps, a frequency roughly the lowest of sound that produces a sensation truly tonal in character. This scale of semitones and fractions thereof can be used for calculation. For example, frequencies 400 and 500 cps are respectively 55.35 and 59.21 semitones above the reference frequency. The difference is 3.86 semitones = 386 cents, the interval corresponding to a frequency ratio of 5/4.

The relatively recent availability of electrical and electronic musical instruments, and the possibility of rapid switching to different temperaments, have served to reopen the question of the tuning of keyboard instruments. Perhaps in another century, the equally tempered scale will have been replaced —but if so, only in satisfaction of requirements of the musical ear and not of a mathematical postulate! See SOUND RECORDING. [R.W.Y.]

Bibliography: J. M. Barbour, *Tuning and Temperament*, 1951; A. Carse, *Musical Wind Instruments*, 1939; C. A. Culver, *Musical Acoustics*, 4th ed., 1956; G. Groves, *Dictionary of Music and Musicians*, 5th ed., 9 vols., 1954; H. F. Olson, *Musical Engineering*, 1952; C. E. Seashore, *Psychology of Music*, 1938.

Musical instruments

Instruments for producing musical sounds for the most part can be classified as woodwinds, brass, percussion, or strings; electrical and electronic instruments, however, are also in use. In a sense, these instruments merely implement and extend the capability of the voice as the original tool by which man expressed himself musically. The categories mentioned are useful for description, even though woodwind instruments are not necessarily made of wood, nor are brass instruments necessarily made of metal. Woodwind instruments are distinguished by the fact that the effective length of a vibrating air column is varied by opening and closing lateral holes. In brass instruments the air is set into vibration directly by the lips of the player; in modern brass instruments the length of the contained air column is modified by switching additional tubing by the use of valves, either piston or rotary. The playing ranges of both classes of instruments are extended by use of different modes of vibration of the air column.

Frequency range. The accompanying chart shows by horizontal bars the playing ranges of many of the musical instruments in common use today. For example, the notes playable on the usual flute go from middle C to the C written above the fifth ledger line above the treble staff. There are numerous notations for expressing such a statement succinctly; by the one employed in the chart, the range of the flute is from C_4 to C_7 inclusive. Notes within an octave ascending from a given C receive the subscript of that C. The order in which instruments are listed in the chart is approximately that employed in an orchestral score. The playing range of the piano is represented by the keyboard at the bottom of the chart.

The ranges shown are for instruments as they are ordinarily made; the flute, for example, occasionally has an extra key to extend its range down to B_3. In cases of transposing instruments (those not in the key of C) the ranges given are for the actual sounds and not for notes written for the instrument: for example, the low E_3 written for the B♭ soprano clarinet corresponds to the D_3 shown on the chart. The mechanism of a musical instrument commonly sets the lower limit to its playing range, whereas the skill of the performer often determines the upper limit.

Woodwind instruments. For the woodwinds listed in the chart, two distinctly different means of generating sound are employed. For the flute and its small version, the piccolo, the player blows across an opening near one end of the tube in such a way that periodic puffs of air enter the tube. This method of excitation leaves the end of the tube essentially open in the sense that the tube functions very much as a simple tubular resonator with the end open to the atmosphere.

For the oboe and bassoon, a pair of thin reeds (pieces of cane appropriately thinned, shaped, and bound together) beat together between the player's lips to change the player's breath to puffs of air entering the vibrating air column within the instrument. For clarinets and saxophones, a single reed attached to the mouthpiece by a ligature functions in a very similar way. The portion of the mouthpiece against which the reed beats must be appropriately curved; the character of the sound is modified somewhat by the shape of the mouthpiece as well as by the shape of the reed. The vibrating reed allows puffs of air to enter when the sound pressure within the instrument is large, and as a consequence (in contrast to the flute) the end at which the mouthpiece is attached is essentially closed. Thus if a clarinet mouthpiece is attached to a flute body the resulting frequency of vibration is about half of that produced with the flute mouthpiece.

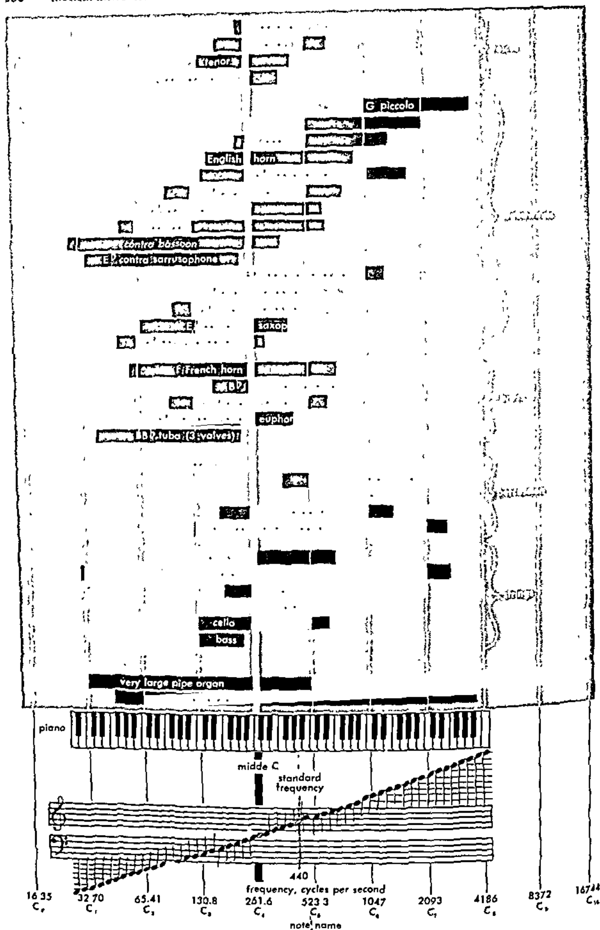
Brass instruments. For the brass wind instruments the puffs of air are introduced to the vibrating air column via the vibrating lips of the player which have been stretched across the cup-shaped opening of the mouthpiece. Again the action of the vibrating lips is such as to close the end of the tube effectively.

The quality of sound from a wind instrument is determined both by the shape of the resonator constituted by the body of the instrument and by the player himself. Woodwind instruments approximate simple geometric shapes: the flute and clarinet are roughly cylindrical, whereas the oboe, saxophone, and bassoon are each based on several cones of slightly differing tapers. In most brass instruments there is a cylindrical section in conjunction with the more obvious expanding shape, so that the over-all shape of the contained air column cannot be described by one simple mathematical curve.

It is noteworthy that the various shapes that have evolved for different wind instruments are all such as to yield nearly harmonic modes of vibration. That is, the frequencies of the different characteristic ways in which the air within the resonator vibrates are nearly in the ratios of small integers. This is not an accident. Most resonators of arbitrary shape do not have harmonic modes of vibration, so the shapes employed for wind musical instruments must be the result of empirical selection by instrument makers over the centuries.

Percussion instruments. Instruments such as the timpani (kettledrums) and xylophone are called percussion instruments because the sound is initiated by a blow. Two kinds of sound producers are involved: a membrane under tension, associated with a hollow cavity that can influence the frequency of vibration, as in the timpano; and a rigid bar vibrating transversely, whose frequency is little affected by any resonator that may be attached, as in the xylophone. Other percussion instruments are bells, drums, cymbals, and triangles.

Stringed instruments. In the case of both the guitar and the harp, strings are set into vibration by plucking. Although a resonator or soundboard



tenor
G piccolo
English horn
contra bassoon
E♭ contra sarrusophone
saxop
F French horn
euphon
BB♭ tuba (3 valves)
cello
bass
very large pipe organ
piano
middle C
standard frequency
440
frequency, cycles per second
16.35 C_0
32.70 C_1
65.41 C_2
130.8 C_3
261.6 C_4
523.3 C_5
1047 C_6
2093 C_7
4186 C_8
8372 C_9
16744 C_{10}
note name

of some kind is attached to help radiate sound to the surrounding air, the frequency of vibration is primarily established by the length, tension, and linear density (mass per unit length) of a string. The same is true of the other stringed instruments listed in the chart, even though the vibration is initiated and maintained by bowing.

Whether a vibration is maintained or allowed to die away naturally is of some importance to the frequencies of the partials which constitute the complex sound usually generated by a musical instrument. Most steadily maintained sounds accepted for musical purposes consist of harmonic partials, whereas the decaying sound of a vibrator such as a piano string or a bell is made up of inharmonic partials. (For definition of harmonic and inharmonic partials, *see* PARTIAL TONE.) If the steel E string of a violin is plucked, the resulting partials are likely to be inharmonic, but if the same string is bowed steadily the partials of the resulting sound are harmonic. Sounds produced by musical instruments are usually complex, consisting of a fundamental and a long series of partials extending in frequency well beyond 10,000 cycles per second.

Keyboard instruments. Instruments such as the celesta, pipe organ, accordion, and piano are sometimes put into a group called keyboard instruments, because the individual vibrators are selected by use of keys in a keyboard. Such classification is convenient because otherwise there are certain complications if one tries to fit them into the groups previously described: the pipe organ keyboard controls pipes that function in the manner of a flute, and sometimes also in the manner of percussion instruments like the xylophone and drum; the freely vibrating reeds of the accordion are in a sense rigid vibrators, but are maintained in vibration by wind; the piano is certainly a percussion instrument as well as a stringed instrument.

Electrical and electronic types. Musical instruments of the kinds already described become quasi-electrical instruments by the addition of a microphone to pick up the airborne sound, an amplifier, and a loudspeaker. Alternately, a vibration pickup can be used to generate an electrical signal directly from the vibration of a string; this is the case in the electric guitar and electric piano. The signal can be modified electrically before being radiated as sound from a loudspeaker. Since the resonance box or sounding board is not needed for these electrical stringed instruments, the damping is very slight and the vibration can last for a long time so that relatively sustained sounds can be obtained if desired.

Rigid vibrators, such as reeds, rods, bars, or tubes, with associated pickups are also exploited as electrical musical instruments. An example is the electronic carillon, in which the vibration of a rod is initiated by a hammer mechanism, and the shape of the rod is so chosen that its modes of vibration simulate in frequency those of the traditional bell. Reeds maintained in vibration by flow of air, as in a reed organ, are found in some electrical organs. The musical quality of the sound from the loudspeaker depends on the location and nature of the pickups.

Rotating disks or drums or moving tapes are used to generate electrical oscillations at the frequencies required for musical purposes by interrupting a light beam which in turn falls on a light-sensitive cell, or by periodically varying the electric or magnetic characteristic of a circuit. The last named method was exploited by T. Cahill about 1900, but its extensive commercial use came with the Hammond organ, on which the first patent was granted in 1934. In this instrument, toothed iron wheels rotating in magnetic fields generate sinusoidal signals at frequencies of the equally tempered chromatic scale; below $F^{\sharp}_8$ and within its playing range upward from C_1, these signals can be combined as quasi-harmonic partials to synthesize a wide variety of waveforms, with amplitudes selected at will.

In the electronic organ a complex signal is first generated by purely electrical means, which usually include an electron tube, and then subjected to various waveform modifications under the control of the player. The onset and decay of sound may also be modified by the player, and these transient parts of the sound can be as significant musically as the steady part. Stops can be combined in the manner of the traditional pipe organ.

Most of the electrical instruments already mentioned are under the control of a keyboard and are capable of sounding several melodies at the same time: thus they are polyphonic instruments. Electrical monophonic (solo) instruments are, however, available, and some afford the same flexibility of intonation as does the violin.

Efforts at artistic combinations of electrically recorded and generated signals promise to open a more direct channel of communication from the composer to the listener. The electronic music synthesizer is a special machine for this purpose; it contains electrical means for producing practically any frequency, growth, duration, decay, portamento, or vibrato under control of a coded record. The development of new musical instruments has consistently inspired new kinds of musical composition. Science has thus far contributed little to the development of traditional musical instruments, but it is now in a position to make a real contribution to this new musical instrument that is played directly by the composer. *See* MUSICAL ACOUSTICS; RESONANCE (ACOUSTICS AND MECHANICS); RESONATOR, ACOUSTIC; SOUND REPRODUCTION SYSTEMS, ELECTRICAL; SYMPATHETIC VIBRATION; VIBRATION.

[R.W.Y.]

Bibliography: *See* MUSICAL ACOUSTICS.

Frequency ranges of music. (Adapted from a chart copyrighted by the W. M. Welch Scientific Co.)

Musk ox

A large Arctic grazing animal, *Ovibos moschatus*, of the family Bovidae, order Artiodactyla. The musk ox somewhat resembles the American bison, but is covered by a dense coat of long, shaggy, brown hair shading to black on the chest; the legs are whitish. The musk ox may reach a length of 8 ft and a weight of 700 lb. This hardy mammal is found only along the northern edge of Greenland, in the extreme north central part of Canada, and on the intervening islands. Although it is nearly extinct, the musk ox has some potential value as a domesticated animal. See ARTIODACTYLA. [J.D.B.]



The musk ox, Ovibos moschatus; length to 8 ft.

Muskmelon

The edible fruit of *Cucumis melo*, a trailing annual of the plant order Campanulales and native to the Near East and Middle Asia. American varieties range in fruit characters as follows: shape, globular to long tapered; weight, 2–9 lb; surface, uniform to broadly sutured to wrinkled, and smooth to heavily netted; color of flesh, pale green to orange; odor and flavor, slight to distinctly musky. The blossom end of the fruit is sweeter than the stem end. The branching vines have coarse, somewhat heart-shaped leaves, with almost entire, slight-angled, rounded, wavy margins. Generally both male and female flowers occur on each plant; in some varieties perfect (combined male and female) flowers are borne. The time from planting to harvest is 85–125 days.

The so-called cantaloupes of American markets are all muskmelons, as are honeydew melons, Persian melons, casaba melons, and such unusual forms as the Banana and Crenshaw varieties. These melons differ only in varietal characters and all intercross freely.

Muskmelons grow well on fertile, well-watered soils, but efficient production is limited by climate. They require fairly warm weather and are favored by bright sunshine and dry air. In rainy or humid areas, downy mildew seriously impairs yields and quality. In some irrigated districts, powdery mildew attacks most varieties. Virus diseases are common. Disease control is a major problem for muskmelon growers. Fruits from seriously diseased vines lack sweetness and characteristic flavor, regardless of their appearance.

Harvesting. With certain exceptions, the stem and the fruit begin to separate as the fruit approaches ripeness. When the separation crack is visible entirely around the stem where it joins the fruit, the stem can be cleanly separated by slight movement without breaking the stem or fruit tissues. At this so-called full-slip stage the fruit contains its maximum of sugar and is ready for harvest. If the stem is removed before full slip, broken tissue of stem or fruit is evident, and the fruit has not reached its best quality. Fruits will soften after harvest, but do not become noticeably sweeter. The winter type of muskmelon, such as honeydew, Persian, casaba, and Crenshaw, do not form a stem-separation layer until after they are ripe. Ripeness in these is judged by color; their fruits are harvested by cutting the stems.

Culture. Except for differences imposed by soil or climate, cultural methods are similar for all kinds of muskmelons. For earliest marketing, the planted seeds (or small transplants) may be covered with sheets of glassine paper measuring 18 by 20 in., over wire arches; or other small paper covers may be used. These remain until the plants become crowded within them or warm weather necessitates removal. Windbreaks are sometimes provided for each row so covered.

California, Arizona, Texas, and Colorado account for about two-thirds of the acreage and three-fourths of the production of muskmelons in the United States. However, muskmelons are grown commercially in about 20 other states, some as far north as New York and Washington. The average annual farm value of muskmelons in the United States during the period 1949–1957 was over $56,000,000. See CAMPANULALES; CANTALOUPE; HONEYDEW MELON; PERSIAN MELON; see also MELON GROWING. [V.R.B.]

Muskrat

An aquatic rodent, *Ondatra zibethica*, with rich, durable, brown fur, belonging to the family Cricetidae. This animal occurs over most of North America north of Mexico. The muskrat has a body length of 10–14 in. and a black, vertically flattened, naked tail 8–11 in. long. It lives either in houses built in



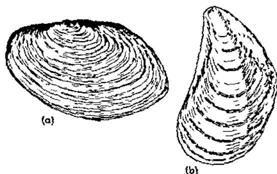
The muskrat (Ondatra zibethica), order Rodentia Length to an average of 22 in.

open water and extending above the water line, or in dens dug into stream banks. The muskrat is most abundant in marsh areas where there is a good supply of rooted aquatic vegetation, which forms its principal food.

Muskrat pelts command a good price on the market. In recent years the muskrat has ranked first among all animals as a source of income for trappers. Approximately 10,000,000 pelts are harvested each year. This rodent is also the staple food of the mink, and the latter seldom thrives except where muskrats are plentiful. *See* RODENTIA. [J.D.B.]

Mussel

A name commonly applied and in this article limited to the fresh-water representatives of the class Pelecypoda, phylum Mollusca. The name clam is also used for the fresh-water bivalves, but is more commonly applied to the marine forms. In like manner, the name mussel is sometimes used for some of the marine forms.



Mussel. (a) Fresh-water, *Lampsilis siliquoides;* shell about 2 by 4 in (b) Marine, *Mytilus edulis;* length to 4 in. (From E. L. Palmer, *Fieldbook of Natural History,* McGraw-Hill, 1949)

Mussels are world-wide in distribution, and occur in all fresh-water habitats, although they are most abundant in large rivers, where they reach their greatest diversity. They range in size from 2 mm to 1 ft in length.

Uses. Mussels are occasionally utilized for food, but not on a commercial basis. They are also commonly used as fish bait and their shells are used to make buttons. There is still a considerable commercial fishery for mussels, especially in the clear, unpolluted streams of the South and Middle West. Arkansas, Kentucky, and Tennessee are the principal states producing shells for buttons. The button industry produces an average of more than $3,000,000 worth of merchandise a year. A few pearls of value are harvested every year, but are becoming increasingly rare as the mussel beds are depleted by overexploitation, pollution, silt, and interference with their distribution by dams which obstruct the migrations of their larval hosts.

The number of species of mussels occurring in the United States is uncertain, but there are at [illegible]

usually alike. The animal moves by protruding the large muscular foot between the valves of the shell and moving over or through the substrate. As in other bivalves, the foot is withdrawn and the shell clamped shut by strong muscles when the animal is disturbed.

The sheet of tissue called the mantle, a fundamental characteristic of the Mollusca, is well developed, lines the shell, and hangs like a curtain on either side of the foot. The visceral mass lies above the foot. Below the middorsal line there is a pericardial cavity containing a three-chambered heart that is connected with anterior and posterior arteries. The circulatory system is of the open type, with a single collecting vein picking up the blood from some of the visceral organs (marine species have paired veins). A pair of kidneys lies below the heart, and below that are the gonads.

The digestive system consists of a pair of thin labial palps which guide food into the mouth, a short esophagus, a prominent rounded stomach, a digestive gland frequently called the liver, and a moderately coiled intestine which passes through the pericardial sac, and the anus which opens into the excurrent siphon.

The nervous system in mussels is quite simple. There are three pairs of ganglia connected by major nerves. Lesser nerves supply the various parts of the body. The cerebral ganglia lie just above the esophagus; the pedal ganglia are well down into the foot; and the visceral ganglia are posterior to the kidneys.

There are two pairs of gills hanging on either side of the foot and inside the mantle. The mantle aids the gills in respiration. Both mantle and gills are covered with cilia which trap food particles as the water, entering the mantle cavity through the posterior and ventral incurrent siphon, passes over them. Food is moved along food grooves on the ventral edge of the gills to the mouth. Water leaves the mantle cavity through the dorsal siphon which is also located posteriorly.

Large, powerful anterior and posterior adductor muscles close the shell. The shell valves are connected with the elastic hinge ligament which tends to draw the dorsal margins together and open the valves, unless opposed by the adductor muscles. The foot is broad and strong, and made up of the large foot muscle.

In addition to its functions in respiration and food gathering, the mantle secretes the shell. The shell is made up of three layers. The outer, horny, and colored periostracum protects the shell from acids in the water which might tend to dissolve it. The bulk of the shell is the prismatic layer, made up of crystalline calcium carbonate, which in some species is relatively thick, in others paper thin. These two layers are secreted by the edge of the mantle. The inner, or nacreous, layer is also secreted

by the edge of the mantle. This iridescent layer is commonly called mother-of-pearl. Pearls are secreted by the same cells when an irritant, such as a grain of sand, is introduced between the mantle and the shell. Organic detritus and plankton are the principal foods of the mussels.

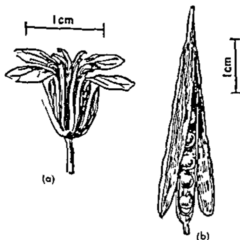
Reproduction. The two families of fresh-water mussels have distinctly different methods of reproduction In the Sphaeriidae, all of which are hermaphroditic, the eggs are fertilized, and the young develop in enlarged segments of the gills called marsupia. From 1 to 20 young develop at a time, but reproduction is probably continued throughout most of the year.

The Unionidae are usually bisexual, but some members are also hermaphroditic. These forms produce as many as 30,000 eggs at a time during late spring and early summer The eggs are fertilized in the gills and undergo their early development there, but are soon shed into the water as minute bivalved animals called glochidia.

The glochidia are parasitic on the gills, fins, or body surface of fish. Some species are highly selective in their host, being monospecific; others can develop in or on a wide variety of fishes. The young glochidium is soon overgrown by the tissues of the host and remains as a parasite for 10–30 days or longer before breaking out and falling to the bottom to continue development as a small mussel. Upstream migration of fishes during spawning runs are an important means of dispersal. This peculiar life history appears especially adapted to maintaining a relatively sedentary group in a running-water habitat, and is in sharp contrast to the pattern exhibited by the marine forms. *See* CLAM; MOLLUSCA; PELECYPODA. [J.D.B.]

Mustard

Any one of a number of annual crucifer species of Asiatic origin belonging to the plant order Papaverales. Mustards eaten as greens are *Brassica juncea*, *B. juncea* var. *crispifolia*, and *B. hirta* Table mustard and oils are obtained from *B. nigra*. Cultural practices are similar to those used for spinach. Southern Giant Curled and Ostrich Plume are popular varieties for greens. Long days and high temperatures favor undesirable seedstalk development. Harvesting is usually 1½–2 months after planting. Important production centers for mustard greens are in the South, where the crop is popular. Montana and the West Coast states are important sources of mustard seed. *See* PAPAVERALES; SPINACH; VEGETABLE GROWING. [H.J.C.]



Mustard. (a) Flower. (b) Pod or silique (fruit) containing seed (From L. H Bailey, ed., The Standard Cyclopedia of Horticulture, vol. 2, Macmillan, 1937)

Mutation

Any inherited change in the genetic material of an organism. A distinction is made between chromosome mutation, or chromosomal aberrations produced by gross alterations in chromosome number or structure, and gene mutations, or internal alterations of the gene itself. This distinction is useful, but it should be realized that there is a spectrum of single chromosome aberrations ranging from rearrangement of almost the entire chromosome to changes so minute as to be hardly discernible at the highest resolution of the microscope. Therefore, certain gene mutations may involve, on a smaller scale, the same qualitative types of change observed in chromosomal aberrations.

From an experimental point of view the distinction between chromosome and gene mutation is not satisfactory, since the criterion of a gene mutation is a negative one. If an inherited change takes place in the genetic material of an organism and by all available techniques is shown to involve neither chromosomal aberration, nor recombination, nor any other known genetic process, then the investigator assumes that a gene mutation has occured. Thus gene mutation is the category into which is lumped those genetic changes which cannot be characterized in other ways.

Mutations may arise in any type of cell carrying inheritable material. If one occurs in a somatic [illegible] mutation to their daughters. If these cells give rise only to somatic tissue, the mutation will be eliminated with the death of the individual. However, if the mutation occurs in a cell which gives rise to germ tissue, that is, in a spermatogonium or an oogonium, or if the mutation takes place in a gamete, then it can pass from one generation to the next. Therefore, germinal mutations form the basis of the inherited differences between individuals.

From this fact, it is obvious that if there were no mutations there would be no genetic differences among organisms and no evolution of life. In fact, mutations have often been called the building blocks of evolution. For this reason, the mutation process is not only fundamental to genetics but to the continuation of life itself in a world that is changing.



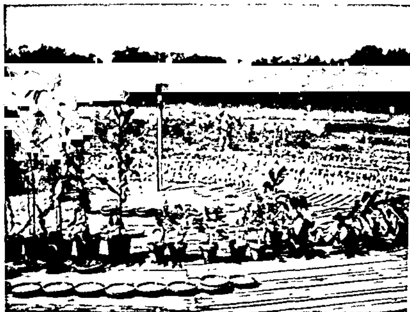
(a) Somatic mutation induced in snapdragon (*Antirrhinum*) after chronic irradiation with γ-rays The plant was heterozygous for flower color and the mutation is from red to white

(b) Carnation with aberrant red flower induced by x-ray treatment of a young plant with 5000 roentgens. Normally only white flowers are produced This appears to be a mutation but actually is not.

(c) Somatic mutation for red to white flower color induced in a tobacco plant (*Nicotiana langsdorffii*) by x-irradiation The plant was exposed to 4000 roentgens at an earlier stage in its development



(d) Somatic mutation (white area in normally all-red flower) in snapdragon (*Antirrhinum*) blossom produced by γ-radiation from the 2000-curie cobalt-60 source in the Brookhaven gamma field. A nonirradiated flower is also shown for comparison



(e) "Atomic farm" at Brookhaven National Laboratory, Upton, N Y., where experiments are conducted on the effects of radiation on plants. When no personnel are in the field, radioactive cobalt raised by the remote control up inside the pipe at center gives off γ-rays which cause genetic changes in plants grown in the field or placed there for varying periods of time. The rays sometimes cause changes which continue to appear in successive generations of a given plant. Some of these changes may prove to be useful

Brookhaven National Laboratory photographs

CHROMOSOME MUTATIONS

Certain of the chromosomal aberrations involve only a rearrangement of the genetic material with no change in the amount of material. The two main types of such rearrangements are inversions and translocations.

Inversions. These are produced by at least two breaks within a chromosome, with a subsequent rotation through 180° of the interstitial broken piece before reunion of the broken chromosome ends occurs. Such a rearrangement does not usually affect the action of genes in the neighborhood of the breaks; it only changes their linear order. However, crossing over between an inverted and a normal chromosome often leads to further chromosome abnormalities such as duplications, deficiencies, chromosome bridges, and fragments which result in nonfunctional gametes or inviable zygotes. Thus, under these conditions this type of chromosome mutation may have rather drastic effects.

Translocations. Translocations usually involve a reciprocal exchange of parts between nonhomologous chromosomes. A simple translocation would require a single break in each of two different chromosomes and a swapping of the two chromosome ends when rejoining takes place. The result is that some of the genes formerly belonging to two distinct linkage groups become linked to each other. Often this produces no effect on the genes concerned, but in a translocation heterozygote (containing one set of translocated chromosomes and one normal set) meiotic segregation produces many gametes having some regions of the translocation duplicated and other regions missing. The end result is that translocation heterozygotes are only partially fertile because of either poorly functioning gametes or inviable zygotes caused by the duplication-deficient gametes which they produce.

Position effect. This is the phenomenon that is the exception to the preceding rule (the action of a gene is not altered by a chromosomal rearrangement when it is placed next to a "foreign" chromosome region) One type of position effect that has often been demonstrated in *Drosophila* and *Oenothera* produces somatic tissue variegated for the expression of the affected gene after it is moved from its normal location in euchromatin to a heterochromatic chromosome region. For example, if an inversion moves the normal allele of the gene pro-

the eyes are not normally pigmented, as expected, but rather are a mosaic of pigmented and nonpigmented, or white, areas. Another type of position effect that is perhaps more general in occurrence is termed position pseudoallelism. In this, the phenotype of the double heterozygote for two closely linked loci depends on whether the mutant alleles of the two loci are in the cis or the trans arrangement. *See* PSEUDOALLELES.

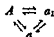
Duplications and deficiencies. These are chromosome mutations which change the amount of parts of chromosomes. A particular chromosome region in a diploid organism may be present three times or only once. These aberrations may arise from breakage of the chromosomes, followed by a pattern of reunion that does not include all the chromatin, or they may be the result of crossing over in inversion heterozygotes or of certain types of segregation in translocation heterozygotes. Since the genes of present-day organisms have had an evolutionary history of selection for harmonious interaction, it is not surprising to find that most of even the very small deficiencies are lethal to the organism when homozygous, and that moderate deficiencies are often lethal, even when heterozygous. Duplications of the same size or region generally have less drastic effects than do deficiencies. There appears to be a balance established through evolution, so that an organism functions most normally when each part of its genetic material is represented the same number of times. Higher plants appear to be much more plastic in this regard than do animals. *See* CHROMOSOME ABERRATION.

Nondisjunction. Nondisjunction is a process whereby a daughter cell may receive a chromosome set differing by one entire chromosome from the normal number Thus, in a diploid organism this process might result in a cell in which one chromosome is represented three times (trisomic) or a cell containing only one representative of this chromosome (monosomic). Nondisjunction arises from a failure of daughter chromosomes to separate at mitotic division or of homologous chromosomes to segregate during meiosis, or from a failure of one chromosome to be included in the daughter nuclei. Since nondisjunction results in duplications or deficiencies for whole chromosomes, it often produces lethality in animals. In higher plants, many trisomic or monosomic individuals do survive and are used extensively in genetic analyses.

Polyploidy. This is the presence of three or more full sets of chromosomes, that is, three or more chromosomes of each and every homologous kind. *See* POLYPLOIDY.

GENE MUTATION

Those alterations in the gene that involve a discrete change from one allelic state to another would seem to form the predominant type of genetic change responsible for the inherited differences between individuals. Often mutations may take place from any allele to any other, as illustrated:



If A is considered the normal or wild-type allele, and the other alleles (a and a_1) are mutant alleles, then the mutation from A to a is called a forward mutation, whereas the change from a to A is called reverse, or back, mutation. One prime characteristic

of a mutant allele is that it is usually just as stable as, and often more stable than, *the wild-type allele from which it was derived.* Another characteristic of a mutant allele is that *it is more likely to be recessive to the wild-type allele than dominant. Finally, most mutant alleles, when expressed, are deleterious to the organism. Some genecticists estimate* that only one mutation in 500 is beneficial. *This* should not be surprising if it is *recalled that* the functioning of the complex parts of an *organism* has been molded into harmony by natural selection and any *randomly produced* change, such as mutation, would more likely *do harm than* benefit.

The changes *in the macromolecular* structure of the gene which form the basis *of gene mutation* are not known. If, as has been suggested by J. D. Watson and F. H. C. *Crick*, the genetic specificity resides in the *sequence of nucleotide* bases in the DNA (deoxyribonucleic acid) *molecule*, then a gene mutation may involve *only the substitution* of one or more bases for those normally *present. It* has also been suggested that a tautomeric shift of one of the hydrogen *atoms* of a base will cause it to pair with a *different complementary* base, thus altering the sequence *in the newly* formed *DNA* chain.

From an *operational* viewpoint, point mutations are those which *are not related* to any detectable chromosome *alteration; or are* not correlated with any phenomenon, like *recombination* or gene conversion. The term point denotes *mutations* that have been defined on an experimental *rather than* a theoretical basis. Special techniques and a genetically well-analysed *organism* are necessary to make critical tests for the *occurrence* of point mutations.

TECHNIQUES FOR STUDYING MUTATIONS

While *chromosome mutations* may be studied directly by cytological *means in* suitable material, their effects, as well as the *effects of* gene mutation, are also studied by genetic means.

Cytological techniques. Chromosome *aberrations*, like broken chromosomes, bridges, translocations, inversions, and *others*, are observable with a reasonable frequency *after* treatment with a mutagen. Only four of the *many organisms* used in this type of study will be mentioned.

The spiderwort, *Tradescantia*, has been one of the higher plants favored for *studies on* the induction of chromosomal aberrations. *Often the* inflorescences are treated with the mutagenic *agent* and, *after a* predetermined period of time, the tissue is fixed and the aberrations *scored* at the metaphase of the microspore division *following* meiosis. The chromosomes of *Tradescantia are* large in size and few in number, and thus *accurate* observations may be made.

The mitotic divisions in the *root tips* of the broad bean, *Vicia faba*, have readily observable chromosomes and since the *root is* easily amenable to chemical treatments that *affect* chromosome breakage and rejoining, it has *been used* extensively in the analysis of chromosome *mutation*.

The giant salivary gland chromosomes of the fruit fly, *Drosophila*, make possible the most critical of the analyses of chromosomal aberrations. The linear, banded appearance of these chromosomes, together with the specific pairing of each corresponding band of homologous chromosomes in somatic synapsis, allows a detailed description of the structure of *rearrangements*.

Finally, it should be mentioned that, beginning *about 1950*, techniques have been developed for chromosome studies in cells grown in tissue culture from many types of vertebrate animals, including man. These techniques hold great promise for determining the effectiveness of various mutagenic agents on human chromosomes. *See* CULTURE, TISSUE.

Genetic techniques. Mutations are, of course, studied not only by cytological methods but also by genetic techniques. Although a genetic method is used, this does not mean that only gene mutations are being observed. For example, inversions and translocations can be detected by their effects on crossing over and inherited semisterility, respectively, without any resort to direct cytological studies.

Sex-linked recessive lethals. These mutations in the X chromosome, extensively studied in *Drosophila*, kill a male receiving that chromosome. They are usually induced by treating the germ cells and observing the results in a later generation after appropriate crosses. Although a given lethal can be localized within a particular chromosome region, in an appreciable proportion of cases a chromosomal aberration producing a position effect is involved. Therefore, recessive lethal mutations involve not only point mutations but also chromosome mutations.

Specific locus mutations. These are often detected by treating wild-type individuals and crossing them to individuals carrying a series of recessive visible mutants. Normally all offspring would be expected to be wild-type in appearance, since the mutants are recessive. However, if a mutation has been induced at one of the loci concerned, then any offspring derived from the gamete carrying the mutation would show that particular phenotype. This technique has been used in such widely different organisms as maize, *Drosophila*, and the mouse.

Dominant lethal mutations. This is another category of mutations that has been extensively investigated

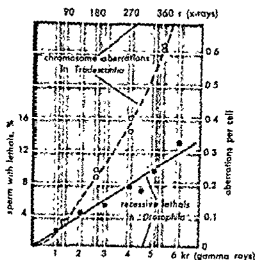
cells

fewer

by chromosomal aberrations and mutations induced in the treated gametes which cause death of the offspring, often during embryonic development. However, in some organisms there is evidence of gamete inactivation and a nonchromosomal basis for lethality.

Mutation studies in man. These include indirect methods, such as sex ratio studies and pedigree analysis, as well as direct methods. Recessive and dominant lethals, if they are sex-linked, affect the

sex ratio. For example, if a man is exposed to radiation, then any sex-linked dominant lethals induced in his X-bearing sperm will cause the death of a female offspring; thus on the average more male than female offspring will be produced. (It is assumed that fewer mutations can occur on the Y than on the X chromosome.) Conversely, if a woman is exposed, the dominant lethals will be distributed equally between her sons and daughters, but the sex-linked recessive lethals will show up in her sons only, so as to result on the average in more daughters than sons. Aside from sex ratio studies of exposed parents, mutation studies in man are limited, in the main, to techniques of pedigree analysis, where, for example, a known dominant condition is seen to occur when neither parent showed the trait, or to indirect techniques involving analyses of gene frequencies and population structure. Tissue culture methods are now making direct measures of human mutation possible. See HUMAN GENETICS.



Relationship between gamma-ray dosage and recessive lethals in *Drosophila*; relationship between x-ray dosage and chromosome aberrations in *Tradescantia*

ORIGIN OF MUTATIONS

Mutations may be produced by various physical and chemical agents. Such changes are induced mutations. Spontaneous mutations likewise occur under natural conditions of the environment, that is, without any artificial application of known mutagenic agents.

Induced mutations. These were first clearly demonstrated by H. J. Muller in 1927, by irradiating *Drosophila* with x-rays. Since that time, extensive studies have been made of the mutagenic effects of other ionizing radiations, including gamma rays, beta particles, fast and thermal neutrons, protons, and alpha particles. By use of specific locus mutation tests with x-rays, it is estimated that the average mutation frequencies in spermatogonial cells of *Drosophila* and mice are 1.5×10^{-8} and 25×10^{-8} per roentgen per locus, respectively. In addition, ultraviolet light and many chemical compounds have been found to be mutagenic.

The relationships between the dose of ionizing radiation, its intensity, its energy, and the frequency with which point mutations and chromosomal aberrations are induced have been studied. In the illustration, the percentages of sperms with lethals is plotted against gamma-ray dose. The resulting curve indicates that sex-linked recessive lethals in *Drosophila* increase linearly with the dose of gamma rays. This relation suggests that there is no dose of radiation, no matter how low, that does *not* have a definite probability of inducing mutation; that is, there is no threshold below which mutation cannot be induced. This fact appears firmly established by both observation and theory.

A second fact concerning the induction of point mutations by ionizing radiations is that the intensity of radiation in terms of the amount of dose per unit of time, over very wide ranges, has no effect on the frequency with which mutations are induced. However, evidence from mouse experiments indicates that a lowering of the dose rate by a factor of 10,000 significantly reduces the number of mutations recovered from irradiated spermatogonia or oocytes but does not affect the mutation frequency of spermatozoa.

Finally, some evidence suggests that the frequency of point mutations may be inversely related to the ion density along the path of the ionizing particle. In Table 1, note the low RBEs for recessive lethal production relative to chromosomal aberrations with fast neutrons and alpha particles.

Target theory. When an ionizing particle traverses matter, its energy is dissipated in the ejection of electrons from atoms, thus resulting in the formation of ion pairs along its path. It is the ionization, and its subsequent chemical effect on the molecule, that causes the biological event (mutation). One of the formal models that has been very fruitful in accounting for the mutagenic action of ionizing radiations is called the target theory; it relates the effects of dose, intensity, and energy upon the induction of point mutations. On this theory, a point mutation is induced by the passage of a single ionizing particle through the cell, because the frequency of mutation increases directly with the

Table 1. Effect of ion density on mutation production

Radiation	Relative biological efficiency (RBE)	
	Chromosome aberrations in *Tradescantia*	Sex-linked recessive lethals in *Drosophila*
Gamma rays from cobalt-60	1	1
X-rays (250 kvp*)	1.5	1–1.6
Fast neutrons	10	1.6
Thermal neutrons	16	?
Internally produced nitrogen protons and boron alphas	22	?
Alpha particles from radon	?	Possibly less than 1

* kvp, kilovolts peak.

number of particles (dose). If a hit by two particles on the target, the gene, were necessary to mutation, then the frequency would increase with a higher power of the dose. Thus one might anticipate that the length of time between two hits (intensity) would modify the frequency of mutation, which is not in general the case. The possible inverse correlation between ion density and the induction of point mutations in certain organisms would be interpreted, according to the target theory, to mean that a single ion pair is sufficient to induce a point mutation. The relative ineffectiveness of the more densely ionizing radiations is caused by wasteful excess of ionizations within the target volume. *See* RADIATION BIOLOGY.

Many types of chromosomal aberrations are formed in chromosome sets in which two or more chromosome breaks have occurred as a result of irradiation. Since with radiations like x-rays or gamma rays the breaks are produced by separate ionizing particles, the individual breakage events are independent of one another. Therefore, the frequency with which these aberrations are induced increases roughly as the square of the dose (see illustration). In many cells the broken chromosome ends stay open for only a short period of time. This results in an intensity effect; if the radiation is administered over an appreciable length of time, then some of the breaks have rejoined before others are formed. The number of aberrations recovered is thus directly correlated with the intensity of radiation. From Table 1 it may be seen that the densely ionizing radiations are more effective, per unit of energy absorbed, in producing aberrations. According to target theory, this means that for chromosome breakage more than a single ion pair must occur within the chromosome volume at a localized site.

Ultraviolet light. The mutagenic effect of ultraviolet light is caused primarily by excitation of atoms rather than ionization. It was once thought that UV produced only point mutations; however, later investigations in *Drosophila* and maize indicate that chromosomal aberrations are induced with appreciable frequency. The action spectrum of ultraviolet light in causing mutation corresponds quite closely to the absorption spectrum of nucleic acids, the maximum absorption and mutagenic action each being about 2650 A (*see* NUCLEIC ACID). This implies that it is the nucleic acid in the genetic material which is undergoing the change.

Modifiers of the effect of ionizing radiations. Further studies have demonstrated that the picture of a particle ionizing the molecules of a chromosome through which it is passing, thus producing a break, is much too simple a reconstruction of the events that actually take place. Various external agents, that have little or no effect on the density of ionization, greatly modify the number of chromosome aberrations recovered. For example, the oxygen concentration within the cells is a potent modifying agent; the lower the oxygen tension, the lower the frequency with which chromosomal rearrangements are recovered. Other effective modifying agents are infrared light and temperature. In some cases these agents appear to change the sensitivity of the chromosome to breakage; in other cases, to modify the ability of the broken chromosome ends to rejoin. It is known that oxidative metabolism is necessary for the reunion of certain broken chromosomes. Mainly these modifying agents are effective only if applied previously or concomitantly with the ionizing radiation. But the mutagenic potency of ultraviolet light may be suppressed by the application of visible light after UV treatment. This process is called photoreactivation. *See* BACTERIAL GENETICS.

Table 2. Some of the known chemical mutagens

Mutagen	Effective in producing	
	Mutations	Chromosomal aberrations
Urethane	+	+
Diepoxide	?	+
Maleic hydrazide	?	+
Diepoxybutane	+	+
Triazine	+	+
Formaldehyde	+	+
Mustard gas	+	+
Nitrogen mustard	+	+
Hydrogen peroxide and organic peroxides	+	+
Caffeine	+	+

Not only is the mutagenic sensitivity of cells altered by artificial application of various agents, but different kinds of cells have differing sensitivities and even the same cell at different stages of differentiation will show varying responses to radiation.

Chemical mutagens. All these secondary factors which influence the effect of radiation point to the possibility that the interaction of protoplasm and ionizing radiation may produce certain chemical compounds or radicals which are in turn mutagenic. During World War II, it was discovered that mustard gas is a potent mutagen. Since that time, a host of chemical mutagens has been uncovered. In Table 2 are listed a few of these compounds which have been extensively studied in so far as point mutations and chromosomal aberrations are concerned. Certain of these chemicals produce primarily point mutations, whereas others induce a relatively high frequency of chromosomal aberrations. In addition, certain of the mutagens break the chromosome at rather specific places and not in a random manner, like ionizing radiations. In fact, in bacteriophages, bacteria, *Neurospora*, and silkworms it has been found that genes mutate with different relative frequencies upon treatment with different chemical agents. The possibility is thus raised that some day it may be possible to direct the mutation process, that is, to find a particular chemical which will induce the mutation of one particular gene only.

Spontaneous mutation. This is a rare event and thus a study of the process is extremely laborious except in situations where a large number of organ-

Table 3. Sample spontaneous mutation frequencies

Organism	Gene	Mutation frequency per locus per generation
Man	Retinoblastoma	2.3×10^{-5}
	Hemophilia	3.2×10^{-5}
	Albinism	2.8×10^{-5}
	Achondroplasia	$1–14.2 \times 10^{-5}$
Mouse	7 loci	0.8×10^{-5} (average)
Drosophila	Yellow	1.2×10^{-4}
	Eyeless	6.0×10^{-5}
	Ebony	2.0×10^{-5}
	5 loci	8.0×10^{-6} (average)
Maize	*R* locus (anthocyanin production)	4.9×10^{-4}
	I locus (colorless aleurone)	1.1×10^{-5}
	Pr locus (purple aleurone)	1.1×10^{-5}
	Sh locus (shrunken endospore)	1.0×10^{-6}
Neurospora	Adenine (reversion)	6.0×10^{-8}
	Inositol (reversion)	$1.5–72.6 \times 10^{-8}$
	Inositol (reversion)	1.0×10^{-8}
Escherichia coli	Streptomycin	1.0×10^{-10}
	T1 phage resistance	1.0×10^{-7}
Bacteriophage T4	Rapid lysis mutants	
	r 101 (reversion)	3.0×10^{-7}
	r 51 (reversion)	1.1×10^{-8}

isms can be readily scored for a change in phenotype. *Table 3 is a list of gene mutation frequencies* covering a variety of organisms and loci. Spontaneous mutation frequencies mainly fall in the range of 1 in 10,000 to 1 in 1,000,000 per generation. Not only do different loci have different mutation frequencies, but also different alleles of the same gene mutate with characteristic frequencies. In fact, there would appear to be as much variation in mutation frequencies in different genes within an organism as there is between organisms. However, part of this variation is undoubtedly caused by differences in experimental techniques. In maize and in *Drosophila* the process of general mutability itself is demonstrably under genetic control inasmuch as the same alleles have different mutation frequencies in different genetic backgrounds.

Although gross chromosomal aberrations must *arise spontaneously, since they form the basis of* differences in the chromosomal karyotypes of many related species, they occur with such a low rate that it has been impossible to obtain any precise figures on frequency. In organisms like bacteria, where extremely low mutation frequencies can be measured, the analytical methods for the detection of chromosomal aberrations are not available.

The basic causes underlying spontaneous mutation are still unknown. Several theories remain tenable. Some theories hold that the natural occurrence of known mutagenic agents might be sufficient to account for spontaneous mutation. It seems certain that natural radioactivity is not present in sufficient quantities to account *completely for the* frequencies observed. There is not sufficient evidence as yet to decide whether naturally occurring chemical mutagens may be significant causes. It has long been recognized that the spontaneous mutation frequency increases with increasing temperature. This observation lends some support to the theory that mutations arise as a consequence of rare "hot spots" that occur when molecules move in a statistically random manner.

Not only are the causes of spontaneous mutation unknown, but also it is not known if these mutations *can take place at all stages of the division cycle of* the cell. One attractive hypothesis is that mutation is an error that takes place while the gene is replicating. This notion is still fruitful as a working hypothesis, since the evidence is contradictory (in bacterial populations) as to whether or not the total mutation frequency is altered when the division rate of the cells is inhibited.

The basis of spontaneous mutation remains one of the great unsolved problems of genetics. *See* GENE; GENETICS. [W.K.B.]

Bibliography: *Mutation*, Brookhaven Symposia in Biol. 8, 1955; A. Hollaender (ed.), *Radiation Biology*, 3 vols., 1954–1956.

Mutation (ontogeny)

A mutation which affects a process of development at any stage prior to that of the fully developed individual, that is, during its embryogeny, is referred to as a mutation in ontogeny. Strictly speaking, *the vast majority of mutations belong in this* category, because every character or trait is subject to the effect of a mutation while it develops, that is, during its ontogeny. Mutations in ontogeny may affect any organism throughout the plant and animal kingdoms and result in arrest or changes of developmental processes. Certain mutations in ontogeny produce large changes of organ character and, according to R. Goldschmidt, are possibly responsible for evolution in large steps, macroevolution. Morphological or physiological abnormalities of the fully developed organism result from the effects of mutations on normal developmental processes. If these effects are severe enough, death during development may ensue. As a result the parents would not contribute to the effective size of the population. *See* MUTATION.

The *study of mutations in ontogeny presents a* number of different interesting facets. To the geneticist their study reveals details of individual steps by which genes produce their effects. The student of development may derive knowledge about steps and mechanisms of normal development from studies of abnormal development caused by mutations in ontogeny. *Changes in developmental processes*, produced by mutations, must precede evolutionary changes of the form and function of adult organs.

Studies. Extensive studies of ontogenetic effects of mutations in plants and in animals such as insects, birds, and mammals have been made by numerous investigators.

The following specific ontogenetic characters of earliest development are affected by single mutations: (1) size and form of chromosomes, (2) condition of chromosomes, (3) chiasma formation dur-

ing meiosis, (4) spindle formation, (5) growth and maturation of oocytes, and (6) polarity and cleavage pattern.

The following mutations have effects during later embryonic life: (1) the mutation *he* which affects the formation of cuticle and the coloration of scales in *Ephestia*, (2) several mutations affecting formation of wing patterns in *Ephestia*, (3) mutations affecting the pattern of bristle formation in the prothorax of *Drosophila*, (4) mutations interfering with normal segmentation of the abdomen in *Drosophila*, and (5) mutations changing the organ character of imaginal disks in *Drosophila*, for example, a change of the arista of the antenna into a leg (aristapedia).

Insects. Some mutations in *Drosophila* with effects during ontogeny severe enough to result in death of the embryo, larva, or pupa include the following:

Mutation "Kruppel" causes abnormalities of organogenesis, for example, defective segmentation, defective tracheal system, abnormalities of internal organs. Homozygotes die as embryos.

Mutation "lethal meander" causes disproportionate growth of larval organs, for example, overgrown tracheal system, overgrown cuticle, but retarded growth of rest of larval body. The syndrome is the result of starvation due to abnormal protein metabolism and results in death of the larvae.

The mutation "cryptocephal" makes it impossible for the head of the insect to be everted at pupation. Inside the puparium the fly is relatively well-differentiated but unable to hatch, and dies in the pupal stage.

Birds. "Frizzle fowl" is an incomplete dominant characteristic in fowl with an immediate effect of mutation on feather development; abnormal feathering responsible for disturbances of body temperature control, increased food consumption, or increased circulation with resulting heart enlargement.

"Wingless" is a recessive mutation in fowl resulting in absence of wings, kidney, and lungs, and abnormal epithelial differentiation of ectodermal, mesodermal, and endodermal derivatives. Homozygotes die before hatching.

"Creeper" is an incomplete dominant characteristic in fowl. Abnormalities of cartilage differentiation in heterozygotes result in typical chondrodystrophy. Most homozygotes die about the third day of incubation. Occasional survivors beyond this stage have severe defects of extremities leading to absence of bone marrow and acute anemia, head and eye abnormalities, retarded growth and death before hatching.

Mammals. Mutations affecting ontogeny of mammals have been studied, particularly in mice. Organ systems of mouse embryos have been shown to be affected by mutations. In the skeleton the following occur: shortened tail, abnormalities of the spinal column, shortened long bones, abnormalities of toe formation, harelip, and cleft palate. In the nervous system, pseudoencephaly and hydroencephaly, disorders of the encephalon, result (*see* TERATOGENESIS). In the hemopoietic system, macrocytic anemias and siderocytic anemia result (*see* HUMAN GENETICS). In the urogenital system the kidneys may be absent or polycystic. In the pigment-forming system, pigment formation is either suppressed or modified.

A considerable number of mutations affect early embryonic processes in the mouse severely enough to cause the death of the embryo. Such mutations have been studied particularly in their effects on the notochord-mesoderm system and are inductive phenomena during embryogeny.

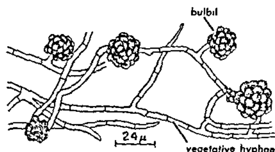
Evolutionary significance. The evolutionary significance of mutations affecting early embryonic differentiation has been studied and analyzed by Goldschmidt in arthropods. Mutations have been shown, for example, to be able to produce abnormalities in specialization of body segments in *Drosophila*; among these are the replacement of antennae by tarsi of the legs, or of eyes by antennae. A single genetic change thus has a large morphogenetic result which can frequently be duplicated by experimental manipulation of the environment. A phenocopy of the mutant aristapedia, for example, may be produced experimentally in *Drosophila* larvae by chemical treatment. Or the genetically produced abnormality may be found in nature as a normal characteristic in other taxonomic groups, such as the transformation of wings into halteres, occurring in *Drosophila* as the result of a mutation and normally in certain Neuroptera and Hymenoptera.

Goldschmidt has expressed the idea that mutations producing monstrosities by affecting ontogeny may have played a significant role in macroevolution. A mutation in fish which disturbs skull development produces a monster. However, the same mutation in the flounder enables it to live on the sandy bottom of the ocean. Achondroplasia as a mutation in dogs affects the development of the bones in such a way as to produce a monster; however, the dachshund with achondroplastic extremities fills a special role in being able to follow the badger. Goldschmidt calls such mutants hopeful monsters and considers them one of the means of macroevolution by single large steps. Further examples for this type of mutation are hairlessness, taillessness, wing rudimentation, reduced eyes, and telescope eyes. Such monstrosities are a result of mutational changes in early embryonic processes. *See* EVOLUTION, ORGANIC; GENETICS; LETHAL GENE; POPULATION GENETICS. [S.C.W.]

Mycelia Sterilia

An order of fungi of the class Fungi Imperfecti. Certain members of the Mycelia Sterilia are pathogens of potatoes, onions, garlic, and other wild and cultivated plants. Mycelia Sterilia may be the imperfect (asexual) stage of certain Ascomycetes, Basidiomycetes, or other Fungi Imperfecti.

The Mycelia Sterilia are a heterogeneous collection of Fungi Imperfecti which have no spores.



Papulospora pisicola. Creeping hyphae with bulbils. (After L. C. Doyer, 1938)

Early mycologists gave names to collections of mycelial structures, such as sclerotia and rhizomorphs, for classification purposes but these names were discontinued when it was seen that the structures did not produce spores.

There are approximately 20 genera with 200 species. The following genera have species which are important plant pathogens.

Sclerotium, the nonreproducing form is a dark, tubercle-like mass of hyphae. *S. cepivorum* causes onion white rot. *S. rolfsii* has brown sclerotia and attacks a number of hosts.

Rhizoctonia, small tubercles are connected by mycelial threads. *Rh. crocorum* causes violet root rot and is the imperfect stage of *Helicobasidium purpureum* (subgroup Basidiomycetes). *Rh. solani* causes "damping off" and is the imperfect stage of *Pellicularia filamentosa* (Basidiomycetes).

Pachyma, large subterranean sclerotium, fibrous, fleshy internally and irregular externally. *P. cocos* are the hard sclerotia (stone fungus) of *Polyporus tuberaster.* An edible fruit body known as tuckahoe or Indian bread is produced when *P. cocos* is watered.

Papulospora are small, rounded clusters (bulbils) of light brown cells resembling small sclerotia. The members of this genus are saprophytic *P byssina* is the brown plaster mold of mushroom beds. See Fungi; Fungi Imperfecti. [N.F.B.]

Mycetozoida

An order of the subphylum Sarcodina commonly termed slime molds by botanists. They are Rhizopodea in which the mature stage is a plasmodium, often a plasmodial net, or, in the Acrasina, the cellular slime molds, a pseudoplasmodium in which there is no protoplasmic fusion. Most are free-living, but the Plasmodiophorina are parasites of plants, often associated with tissue hypertrophy and galls such as the club root of cabbage. The plasmodial Mycetozoida (Eumycetozoina) usually occur on decaying plant debris and manure. At maturity, the plasmodium may measure several inches and often contains pigment which may be violet, blue, green, yellow, orange, red, or brown. Holozoic feeding is common, but all species are probably saprozoic.

Reproduction. Reproduction in the Eumycetozoina involves sporulation (Fig. 1). Resistant spores are produced, often in sporangia, the process resembling that in many fungi. The plasmodium, or portions of it in certain species, when subjected to drying becomes enclosed in a resistant membrane, the sclerotium, and remains viable for months. With adequate moisture, spores germinate (Fig. 2) liberating myxamebae which may give rise to myxoflagellates. The myxoflagellates are believed to be gametes which can undergo syngamy to form ameboid zygotes (see Reproduction, animal). Growth and nuclear division result in a small plasmodium. Continued growth, apparently sometimes augmented by fusion of plasmodia, results in a mature plasmodium which eventually undergoes sporulation.

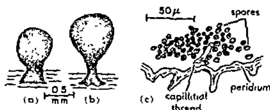


Fig. 1. (a, b) Development of sporangium in the slime mold *Arcyria cinerea* (after Kränzlin). (c) Section of mature sporangium (From R. P. Hall, Protozoology, Prentice-Hall, 1953)

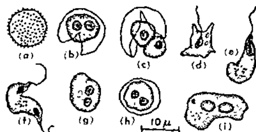


Fig. 2. Stages in cycle of *Physarum polycephalum,* one of the Eumycetozoina. (a) Spore. (b, c) Hatching of spore. (d, e) Myxoflagellates. (f, g, h) Zygote before nuclear fusion. (i) Excysted zygote after first nuclear division (after Howard). (From R. P. Hall, Protozoology, Prentice-Hall, 1953)

In the Acrasina, uninucleate amebas migrate to a focal point and adhere to form a pseudoplasmodium. This aggregation is a response to a specific substance, "acrasin," secreted by certain amebas, called "evocators." The pseudoplasmodium, in *Dictyostelium,* for example, migrates as a unit, although protoplasmic fusion does not occur. It eventually differentiates into a pseudosporangium, composed of a base, stalk, and spores. See Myxomycetes; Rhizopodea. [R.P.H.]

Mycobacteriaceae

A group of nonsporeforming, nonmotile, aerobic gram-positive bacteria belonging to the order Actinomycetales. The genus *Mycobacterium* forms the

main group and comprises bacteria with irregularly rod-shaped cells that resist decolorization with dilute inorganic acids after staining with anilin dyes (acid-fastness). Growth in artificial media is often slow, with little or no acid formation from carbohydrates. The genus includes some important pathogens, for example, *M. tuberculosis* which causes tuberculosis in man and warm-blooded animals, and *M. leprae* which causes leprosy; others cause tuberculosis-like lesions in cold-blooded animals. Saprophytic species, for example, *M. phlei*, are common in soil and dung. Several of these utilize hydrocarbons as energy source. Some of them are only slightly acid-fast, tend to form branching cells, and are thus difficult to distinguish from the genus *Nocardia* See ACTINOMYCETACEAE.

A genus *Mycococcus*, consisting of soil bacteria with very short, almost spherical cells arranged in short branched chains, is included in the Mycobacteriaceae, but little is known about its real nature. *See* ACTINOMYCETALES; LEPROSY; SOIL MICROBIOLOGY; TUBERCULOSIS. [H.L.J.]

Mycology

A field of microbiology concerned with the study of fungi, those microorganisms included in the division Thallophyta of the plant kingdom. Fungi lack photosynthetic pigments and include such familiar plants as the molds, yeasts, and bacteria. *See* EUMYCETES; THALLOPHYTA.

Medical mycology is a specialized field of study concerned with fungi that cause disease Practically all of the pathogenic fungi affecting humans belong to either the Schizomycetes or the Fungi Imperfecti; a few species are Phycomycetes.

From a clinical viewpoint two major categories of fungus disease are recognized: those which cause superficial skin infections, the dermatophytes, and those which may produce deep invasive infections. *See* FUNGI IMPERFECTI; MYCOLOGY, MEDICAL; SCHIZOMYCETES.

Extensive investigation into the nature of fungi in recent years has furnished information of value to industry and medicine. Perhaps the most dramatic phase of this investigation has been the development of substances (antibiotics) which have an antimicrobial action. These antibiotics are *bactericidal*, *killing* the offending bacteria, or bacteriostatic, immobilizing them and preventing reproduction. Substances which have a similar effect on fungus infections are also known. *See* ANTIBIOTIC; ANTIMICROBIAL AGENTS; FUNGISTAT AND FUNGICIDE.

The development of effective therapeutic measures for many bacterial diseases has greatly reduced morbidity and mortality. The advent of antibiotic therapy and its wide-scale use has, however, increased the relative importance of fungus infections. Use of antibiotics may suppress the normal bacterial inhabitants of the body, permitting unopposed growth of fungi which may be present (fungi are not affected by many antibiotics). [E.C.ST.]

Mycology, medical

The study of the fungi (actinomycetes, molds, and yeast) pathogenic to man.

Most of those fungi infecting man have been found to exist in nature as saprophytes, and it is believed that their pathogenic properties for man and animals are incidental. *See* FUNGI.

Mycotic diseases are usually differentiated on the basis of the area of the body affected or according to the name of the particular etiological agent. For example, the ringworm infections are known as dermatophytoses. Going one step further, ringworm of the scalp is *Tinea capitis*; of the glabrous skin, it is *Tinea corporis*. Diseases due to *Histoplasma capsulatum* and *Blastomyces dermatitidis* are known as histoplasmosis and North American blastomycosis, respectively.

Fungi are part of the plant kingdom belonging to the phylum Thallophyta. Algae and lichens are also included in this phylum. Fungi do not possess chlorophyll, roots, stems, or leaves. They are broken down into three subdivisions: Schizomycetes (bacteria and actinomycetes), Myxomycetes (slime molds), and Eumycetes (true fungi). The Eumycetes consist of four large groups: Phycomycetes, Ascomycetes, Basidiomycetes, and Fungi Imperfecti. *See* EUMYCETES; MYXOMYCETES; SCHIZOMYCETES.

The simplest of the fungi is the yeast, a unicellular organism. It may reproduce by a simple budding process (asexual) or by development of spores within, as a result of fusion of two nuclei-producing haploid nuclei (sexual spores). Other fungi consist of a thallus which is made up of aerial and vegetative mycelia. Each mycelium consists of many hyphae (filaments) which may be septate or nonseptate. Their classification is based upon type of growth, type of spores produced and how they are produced, and absence of hyphae.

Most of the fungi causing mycotic diseases in man are found in the Fungi Imperfecti; an exception to this are the *Mucor* and *Rhizopus* species, which are classified in the Phycomycetes. *See* ASCOMYCETES; BASIDIOMYCETES; FUNGI IMPERFECTI; PHYCOMYCETES.

Most of the fungi causing infections can be grown on artificial media in the laboratory. The most commonly used of these media is a modification of Sabouraud's peptone-sugar agar. However, this medium is not nutritious enough for the isolation of some of the more fastidious fungi. Isolation of *Histoplasma capsulatum* from tissue is enhanced when an enriched medium such as brain-heart infusion or trypticase soy is used; *Trichophyton verrucosum* requires the presence of thiamine for satisfactory growth.

The addition of such antibiotics as streptomycin, penicillin, and chloramphenicol to media enhances the isolation of pathogenic fungi from materials contaminated with bacteria. Fungi are not affected by these antibiotics, but their growth may be inhibited if large numbers of bacteria are also pres-

ent on the media. *See* ACTINOMYCOSIS; ASPERGILLOSIS; BLASTOMYCOSIS; CANDIDIASIS; CHLORAMPHENICOL; CHROMOMYCOSIS; COCCIDIOIDOMYCOSIS; DERMATOPHYTOSIS; HISTOPLASMOSIS; PENICILLIN; NOCARDIOSIS; STREPTOMYCIN. [L.D.H.]

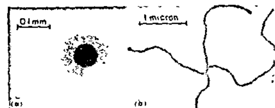
Mycoplasmatales

An order of bacteria of the class Schizomycetes. The microorganisms are better known as pleuropneumonia-like organisms (PPLO). The order includes one family, Mycoplasmataceae, and one genus, *Mycoplasma*. Fifteen species have been recognized to date. The organisms are gram-negative, nonmotile, nonsporing; serum is generally required for growth. They occur free in nature, and as saprophytes or pathogens in animals and man.

Morphology. The microorganisms have filterable elementary bodies, 0.125–0.250 μ in diameter, which germinate to form branching filaments varying in length from 2 to 150 μ. The filaments are subsequently transformed into chains of spherical corpuscles which, upon fragmentation, are liberated as new elementary bodies. The lack of a rigid cell wall is responsible for the frequent development of pleomorphic structures called large bodies Colonies on agar are extremely small; they consist of an opaque, yellowish central area which grows down into the medium, and a translucent, flat peripheral zone. The pleuropneumonia-like organisms and L forms of other bacteria have much in common. Some investigators consider them as growing protoplasts of other bacteria which have become more or less stabilized.

Biochemical characteristics. Some species of this bacterial order ferment carbohydrates. Cholesterol and other lipids play an important role in the metabolism; lipases (enzymes which split lipids) are probably involved in the development of characteristic precipitates formed by certain species on horse-serum agar. Gelatin and coagulated horse serum are rarely liquefied. Alpha hemolysis is sometimes produced on horse-blood agar (*see* BLOOD PLATE HEMOLYSIS). All species are highly resistant to antibiotics like penicillin, erythromycin, and to sulfa drugs such as sulfathiazole.

Two varieties of the type species, *Mycoplasma mycoides*, are the etiological agents of contagious pleuropneumonia of cattle and goats, respectively. *M. agalactiae* causes contagious agalactia, or mastitis, in sheep and goats. A generalized infection



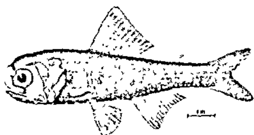
(a) Mycoplasma hominis colony on solid nutrient medium, 5 days old. (b) Electron micrograph of Mycoplasma mycoides var. mycoides from a 24-hour broth culture.

in swine known as Glasser's disease is caused by *M. hyorhinis*. At least one species is etiologically implicated in chronic respiratory disease of chicken and turkey sinusitis. *M. hominis*, *M. fermentans*, and *M. salivarium* are common saprophytes of the mucous membranes of the lower genital tract and oral cavity in man; the first two are potentially pathogenic. A number of saprophytic and pathogenic species occur in mice and rats and may cause arthritis or symptoms related to the central nervous system. *M. laidlawii* is frequently found in soil, manure, and sewage. *See* ANTIBIOTIC; BACTERIA, TAXONOMY OF; PLEUROPNEUMONIA-LIKE ORGANISM (PPLO); SCHIZOMYCETES. [E.A.F.]

Bibliography: R. S. Breed, E. G. D. Murray, and N. R. Smith, *Bergey's Manual of Determinative Bacteriology*, 7th ed., 1957; E. A. Freundt, *The Mycoplasmataceae*, 1958.

Myctophiformes

An order of soft-rayed actinopterygian fishes that includes the lantern fishes, the lizard fishes, and their allies. The myctophiforms, sometimes called Iniomi, differ from the centralized order of soft-rayed fishes, the Clupeiformes, chiefly in that the



Lantern fish, Aethoprora metopoclampa. (After G. B. Goode and T. H Bean, Oceanic Ichthyology, Mem. Museum Comp. Zool., 22 86–87, U.S. Natl. Museum Spec. Bull. 2, 1895)

upper jaw is bordered only by the premaxillae and the mesocoracoid bone is absent Other characters include the orbitosphenoid bone which may be present or absent; the pelvic fin which is abdominal, with 6–13 rays; an adipose fin which is commonly present; the swimbladder which has a duct; and the absence of the Weberian apparatus.

The myctophiform fishes appeared first in the Upper Cretaceous. There are 2 suborders, 10 Recent families, about 56 genera, and nearly 300 species. All are marine; some inhabit beaches but the majority are oceanic. Those that live at moderate to considerable depths commonly have luminescent photophores on the head and body. *See* ACTINOPTERYGII; PHOTOPHORE GLAND. [R.M.B.]

Myiasis

Commonly, the infestation of vertebrates by the larvae, or maggots, of numerous species of flies. Some invertebrates, such as spiders, may be attacked by species of Sarcophagidae, the fleshflies.

These larvae may invade different parts of the bodies of these animals or may appear externally.

In cutaneous myiasis, the larvae are found in or under the skin. There may be a migration of some species of these larvae through host tissues, resulting in a swelling with intense itching. Such a condition is known as larva migrans, or creeping eruption, and may require surgical treatment.

Intestinal myiasis in man is usually the result of accidentally swallowing the eggs or larvae of these flies. It occurs commonly in many herbivores who ingest the eggs when feeding on contaminated herbage. The larvae settle in the stomach or intestinal tract of the animal host.

Cavity, or wound, myiasis occurs when the larvae invade natural orifices, such as the nasopharynx, vulva, and sinuses, or artificial openings such as wounds. A study of wound myiasis led to the use of the larvae of *Lucilia sericata* and *Phormia regina* in treating osteomyelitis and suppurating wounds. These maggots commonly feed on necrotic tissue; there is a danger, however, that normal tissue may be attacked. Maggots used for this purpose are reared under sterile conditions. In the study of the use of maggots in medicine, it was discovered that allantoin, one of the compounds found in the urine of the maggots, had a healing effect on ulcers.

External myiasis includes infestation by those maggots which are blood-feeders.

Of considerable economic importance is myiasis of such domestic animals as horses, sheep, or cat-

Some Diptera causing myiasis

Common name	Scientific name	Host and site of infestation
Muscidae (houseflies)		
Houseflies	*Musca domestica*	Man, gastrointestinal tract
	Muscina stabulans	Numerous animals, intestinal tract
Lesser housefly	*Fannia canicularis*	Man, intestinal tract
Latrine fly	*Fannia scalaris*	Man, intestinal
Sarcophagidae (fleshflies, blowflies, scavenger flies)		
Gray blowfly	*Sarcophaga haemorrhoidalis*	Man and numerous animals, intestinal tract, cutaneous and genitourinary tract
Gray blowfly	*Sarcophaga carnaria*	Man and numerous animals; intestinal tract, cutaneous and genitourinary tract
Fleshfly	*Wohlfahrtia vigil*	Skin of infants; cutaneous
Fox maggot	*Wohlfahrtia opaca*	Man, fox, and mink, cutaneous
Old World screwworm	*Wohlfahrtia magnifica*	Eyes, ears, and nose of man and animals, cutaneous
Tumbu fly, ver du cayor	*Cordylobia anthropophaga*	Man, domestic and wild animals; cutaneous
Congo floor maggot	*Auchmeromyia luteola*	Man; suck blood
Calliphoridae (blowflies)		
American screwworm	*Callitroga americana*	Man, cutaneous
Secondary screwworm	*Callitroga macellaria*	Sheep and goat; cutaneous
Old World screwworm	*Chrysomyia bezziana*	Man and animals; cutaneous
	Chrysomyia megacephala	Sheep, cutaneous
	Phoenicia cuprina	"Strike" of sheep; cutaneous
Surgical maggot	*Phoenicia sericata*	Man, sheep, and goat; wounds
Bluebottle fly	*Calliphora vomitoria*	
Greenbottle fly, surgical maggot	*Lucilia caesar*	Man; wounds
Blackbottle fly, fleece worm	*Phormia regina*	Sheep; cutaneous
None	*Apaulina* sp	Birds, especially nestlings; suck blood
Gastrophilidae (horse bots)		
Botfly	*Gastrophilus intestinalis*	Stomach bot of horse
Nose fly	*Gastrophilus haemorrhoidalis*	Stomach and duodenum of horse
Chin fly	*Gastrophilus nasalis*	Mouth, throat of horse
None	*Gastrophilus inermis*	Rectum of horse
None	*Gastrophilus pecorum*	Pharynx, stomach, and rectum of horse
Cuterebridae (robust botflies)		
Gusano del monte	*Dermatobia hominis*	Cattle and man; cutaneous
None	*Cuterebra* spp	Rodents and small mammals
None	*Bogeria* sp	Rodents and small mammals
Oestridae (botflies, gadflies, nose flies)		
Sheep botfly	*Oestrus ovis*	Sometimes eyes of sheep and antelope; cutaneous
Nose fly	*Rhinoestrus purpureus*	Horse, hippopotamus, nose and eyes of man
Warble fly	*Hypoderma lineata*	Warbles of cattle and reindeer
None	*Hypoderma bovis*	Warbles of cattle and reindeer
None	*Hypoderma crossi*	Goat, India; subcutaneous
None	*Cephalopina titillator*	Camel; nostrils and nasopharynx
None	*Kirkioestrus* sp.	African antelope; nasal passages
None	*Gadoelstia* sp.	African antelope; nasal passages
None	*Cephalopsis* sp.	Camel; nasal passages
Head botfly	*Cephenemyia* spp	Deer; head bots
Stratiomyidae (soldier flies)		
Soldier fly	*Hermetia illucens*	Recorded from intestine of man
Syrphidae (flower flies)		
Rat-tailed maggot	*Eristalis tenax*	Recorded from stomach of man

tle. Cattle afflicted with cutaneous myiasis produce poor-grade hides full of small perforations. The total annual loss in hides is estimated at more than $2,000,000. Heavy infestations of cattle by screwworms result in the death of the host if not treated. Horses afflicted with the stomach bot become emaciated and may succumb to the ravages of these larvae.

The Diptera of medical and veterinary importance are largely confined to the families Oestridae, Calliphoridae, and Sarcophagidae. An interesting ecological phenomenon has been observed in the life cycle of the human botfly. The female oviposits on the undersurface of the abdomen of mosquitoes as they emerge from their aquatic habitats. When the mosquito takes a blood meal from the vertebrate host, the eggs are in close contact with the skin of the host. Sufficient heat from the host's body causes the larvae to hatch from the eggs and these tiny larvae then burrow into the skin and cause a creeping eruption.

Average annual loss due to screwworms and blowflies, affecting cattle, horses, hogs, sheep, and goats during the period of 1940–1944, was $15,000,000. [C.R.C.]

Bibliography: R. W. Hegner, F. M. Root, D. L. Augustine, and C. G. Huff, *Parasitology*, 1938; *Insects*, USDA Yearbook of Agriculture, 1952.



Mylonites. (a) Strained and broken coarse crystals (porphyroclasts) of feldspar and a train of garnet granules set in fine-grained schistose matrix of quartz and feldspar veined with granoblastic quartz. San Gabriel Mountains, California. Diameter, 5 mm. (b) Granite mylonite. Coarse, strained, partially granulated crystals are of plagioclase, microcline, and quartz. The granulated matrix is composed of quartz, feldspar, and biotite. San Gabriel Mountains, California. Diameter, 5 mm. (c) Mylonitic augen gneiss. Ovoid relic crystals of plagioclase and of potash feldspar, in a matrix of muscovite, chlorite, and quartz, traversed by swarms of stringers of later undeformed quartz. Deadman Lake, British Columbia. Diameter, 6 mm (From H. Williams, F. J. Turner and C. M. Gilbert, Petrography, An Introduction to the Study of Rocks in Thin Sections, Freeman, 1954)

Mylonite

A rock that has suffered extreme mechanical deformation and granulation, but has remained chemically unaltered. In the early stages of this process, cataclasites or flaser rocks are produced. These rocks are characterized by the fact that the nature of the parent rock is easily recognized. The mylonites, however, are altered beyond recognition. But in spite of their pulverized condition they are hard, coherent, often black and glassy-looking rocks, because the crushing took place under confining pressures so high that the comminuted particles have been welded tightly together. Pseudotachylite is the ultimate product of mechanical grinding. It is black, glassy, and partly isotropic when viewed under the microscope. X-ray analyses show that it consists of submicroscopically comminuted dust of various minerals, mostly quartz and feldspar. *See* FLASER ROCK; METAMORPHIC ROCKS.

True mylonites still contain relics of uncrushed parent rock in lensoid patches parallel to the direction of movement. But microscopically the minerals of these patches of relic material exhibit obvious strain effects, such as marginal bending and granulation, cracking, and undulatory extinction. Quartz in particular exhibits undulatory extinction, often in combination with the so-called Boehm lamellae. The fine lamellae probably are produced by intracrystalline gliding and are often emphasized by dusty inclusions which are sub-parallel to the basal plane {0001}. Such quartz grains, as well as feldspars, often appear as insets or eyes in the granulated matrix. This has been denoted as porphyroclastic texture.

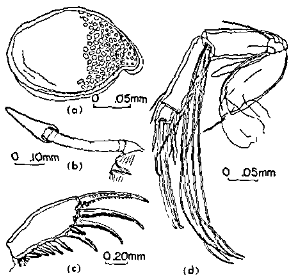
Quartz and feldspar (alkali feldspar and acid plagioclase) are the chief minerals in the most common types of mylonite, because they are brittle and chemically stable over a wide range of temperatures. Thus belts of mylonite in granite, gneiss, and quartzites are common. Mylonites are found in practically all kinds of rock. Worthy of special mention are the dunite mylonites found in many parts of the world.

Laminated fabric is frequently found in mylonites; it can be correlated with flow movements (or creep movements) in the rock during its deformation. A breccia structure also may be present, that is the mylonite mass is made up of angular fragments. The fragments vary in their development according to whether the recrystallization occurred during or after the period of granulation. Rocks which have recrystallized after the granulation are called blastomylonites. [T.F.W.B.]

Myodocopa

A suborder of the class Ostracoda. All members are marine and include both pelagic and bottom species. The shell has a permanent, anterior aperture through which the second antennae may be extended for use in locomotion when the shell is otherwise tightly closed. A well-developed heart is present. The intestine is without cecal diverticula. A frontal organ, located on the head, is characteristic of the suborder.

The paired appendages are the first antennae, second antennae, mandibles, maxillae, three pairs of thoracic legs, and the lamelliform caudal rami, or furcae. The second antennae are the only organs of locomotion in the Myodocopa, although it i



(a) Philomedes longiseta Juday, lateral view of right valve of female *(b) Philomedes carcharodonta* Smith, frontal organ of male. (c) *Philomedes carcharodonta* Smith, furca (caudal ramus) of female. (d) *Pontocypris clemensi* Smith, second antenna of female.

actually the exopodite of each antenna that is provided with locomotor structures.

In the genus *Asterope* Phillipi, sexual dimorphism is very strongly marked in the structure of the antennae and of the shell. The endopodites of the antennae of males are modified as grasping organs.

Some genera of the subfamily Asteropinae can be distinguished from all other ostracods by the presence of true gills similar in structure to those of the higher ...

[E.F.]

Fossil forms. Fossil Myodocopa (Devonian-Recent) are sparse in the geologic record. Fossil ostracods of this order, like their living descendants, have a permanent opening between the valves. *See* OSTRACODA; *see also* SEXUAL DIMORPHISM. [R.V.K.]

Bibliography: G. O. Sars, *An Account of the Crustacea of Norway*, vol. 9, 1928; V. Z. Smith, Further Ostracoda of the Vancouver Island region, *J. Fisheries Research Board, Can.* 9(1):16–41, 1952.

Myriangiales

An order of parasitic fungi of the class Ascomycetes which produce their asci at various levels in stromata which are masses of tightly interwoven somatic hyphae. The stroma weathers away and exposes the asci which contain the spores. These are flask-shaped (globose) and bitunicate, that is, the ascus wall consists of two layers. The elastic inner layer bursts through the firm outer layer and expands greatly, thus elevating the spores which are eventually released through the ascal pore. Examples of Myriangiales are *Elsinoe veneta* and *E. fawcetii*, which cause raspberry and citrus anthracnose, respectively, and *Myriangium duriaei*, which parasitizes scale insects. *Piedraia hortai*, which attacks human hair causing black piedra, has also been placed here. *See* ASCOMYCETES; FUNGI; MYCOLOGY, MEDICAL; PLANT DISEASE.

[C.J.A.]

Myriapoda

Originally, the name devised for an insect subgroup by Pierre Latreille to include the centipedes, millipedes, and isopod Crustacea. By the 1830s Jean Lamarck had elevated Myriapoda to class rank and had restricted it to include only centipedes and millipedes. By 1890 R. Pocock, sensing its unnaturalness, had assigned to each of its chief groups the rank of an independent class, an interpretation still followed today. Within the mandibulate Arthropoda, the myriapodous classes Pauropoda, Symphyla, Diplopoda, and Chilopoda form with the Insecta a group characterized by possession of a single pair of antennae, which are preoral and never leglike, and by their lack of pretarsal levator muscles. By contrast, the Crustacea have two pairs of antennae, the second being postoral and basically leglike, and pretarsal depressor as well as levator muscles.

Myriapoda, then, persists as an informal designation of convenience. It includes those mandibulate Arthropoda having 2 body tagmata, 1 pair of antennae, and more than 3 pairs of adult ambulatory appendages. It is no longer generally believed to delineate a monophyletic assemblage, because, among other reasons, centipedes and symphylans seem more closely related to each other and to lower insects than to diplopods and pauropods.

At the same time the myriapod classes do exhibit a number of commonly shared characters. Excluding two chilopod orders, all are fundamentally anamorphic. All are bitagmate, the body consisting of a head and multisegmental trunk. Most of the trunk segments bear 1 pair of legs each, and in nearly all the leg consists of 7 articles or segments including the pretarsus. In all but Chilopoda the genital ducts open on the anterior part of the body; in centipedes they are posterior and terminal. Until recently it was believed that this distinction between the progoneate classes and the opisthogoneate Chilopoda was indicative of a profound evolutionary divergence; however, recent evidence shows that at least in some Progoneata the anterior ducts are exoskeletal invaginations, not true coelomoducts, and have arisen secondarily. This might mean that the progoneate condition has evolved several times and independently in different groups. Consequently, this characteristic is no longer accorded the evolutionary significance that it once had. *See* CHILOPODA; DIPLOPODA; MANDIBULATA; PAUROPODA; SYMPHYLA. [R.E.C.]

Bibliography: C. Attems, *Myriapoda*, in W. G. ... *Handbuch der Zoolo-* ... off- ... *ca.*

Bull. U.S. Natl. Museum, 212, 1958; J. L. Cloudsley-Thompson, *Spiders, Scorpions, Centipedes and Mites*, 1958; R. E. Snodgrass, *A Textbook of Arthropod Anatomy*, 1952.

Myricales

An order of the plant subclass Dicotyledoneae including the bayberry family (Myricaceae) with 2 genera and 40 species which grow in tropical and temperate regions. They are aromatic shrubs and trees bearing naked unisexual flowers in short scaly aments or catkins. The fruit is a small drupe which, in the wax myrtle (*Myrica cerifera*) and in the bayberry (*Myrica carolinensis*), both native to the eastern United States, has a thick coating of wax used in making bayberry candles. See DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM.

[P.D.S.]

Myrtales

An order of the plant subclass Dicotyledoneae including 12 families, having 380 genera and 9710 species. Taxonomists have variously redistributed these families and there is little agreement concerning relationships. Many of the species are cultivated as ornamentals such as daphne, oleaster, buffaloberry, loosestrife, cigarflower, crapemyrtle, the classical myrtle, tibouchina, evening-primrose, fuchsia, and others. Economically important members are *Lawsonia inermis*, the source of henna; pomegranate; Brazilnut; bay tree (*Pimenta racemosa*), source of an oil used in making bay rum; *P. officinalis*, source of allspice; *Eugenia caryophyllata*, the dried flower-buds of which are commercial cloves; eucalyptus, source of eucalyptus oil, and also a valuable timber tree; the cajeput tree, source of cajeput oil, a universal remedy in the Orient; and guava, an important fruit of tropical America. Here also belong the mangroves occurring in the tropics and subtropics along coasts, shores, and river banks washed by brackish or salt water. See ALLSPICE; BRAZIL NUT; CLOVE; EUCALYPTUS; POMEGRANATE; *see also* DICOTYLEDONEAE; EMBRYOPHYTA; PLANT KINGDOM. [P.D.S.]

Mysidacea

An order of free-swimming Crustacea classified with Cumacea, Isopoda, Tanaidacea, and Amphipoda in the division Peracarida of Malacostraca. They closely retain the caridoid facies. The former association with the Euphausiacea, as "Schizopoda," based on superficial resemblance and biramous appendages, is now obsolete. This order contains two suborders, the Lophogastrida and Mysida. Most species are marine, many are euryhaline and some are fresh-water.

The adult body consists of 19 somites; 5 head, 8 thoracic, and 6 abdominal with a tail fan composed of uropods and a telson. Each somite bears one pair of functionally modified, biramous appendages. The carapace envelops most of the thorax and is fused dorsally with not more than three of the anterior thoracic somites. The eyes are pedunculate and movable. Development is direct and occurs within a ventral brood sac of interlocking lamellae, formed by thoracic appendages of the female. The statocyst is located on the uropod except in the Lophogastrida and primitive Mysida (Fig. 1).

Physiology. Locomotion is effected by a paddling action of the pleopods in lower forms. In higher groups, thoracic exopods set up backwardly directed currents which propel the animal forward. The tail fan and abdomen assist in swimming and can produce springing movements, the "escape mechanism." Walking is accomplished by the thoracic endopods.

A lamellar epipod from each first thoracic appendage maintains currents through a branchial chamber lying between the carapace and thorax on each side. In Lophogastrida the current flows backward over foliaceous gills from the second to the eighth thoracic appendages. In Mysida, it flows forward, generated by thoracic exopods, and gases are exchanged through the thin-walled lining of the carapace, which is richly supplied with blood sinuses Feeding is raptorial in the Lophogastrida, except in the genus *Gnathophausia* which retains the primitive maxillary filter feeding. Living and dead food is captured and held to the mandibles by the subchelate thoracic endopods and mandibular palps and is torn up and chewed. In Mysida, rotatory movements by the thoracic exopods generate currents which, aided by maxillary suction-pump action, flow forward along the ventral food groove. The currents are filtered by setiferous mouth appendages and the resulting food is transferred to the mouth. Filter feeding is supplemented in all bottom-living species by scavenging organic matter from the sea floor.

Sense organs include the eyes and statocysts. Normally the eyes, with movable stalks, bear many-faceted corneas. In abyssal and cavernicolous species, living in total darkness, eyes usually degenerate to simple plates or vestiges lacking visual elements. The statocyst is present at the anterior end of the uropodal endopod in all Mysida except the most primitive families. This is the most conspicuous character of the suborder.

Growth and reproduction. Growth continues throughout life and always follows molting, which occurs at night. In warm temperatures intervals be-

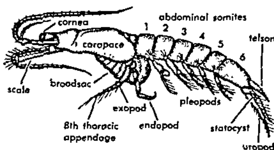


Fig. 1. Diagram of a mysid showing caridoid facies.

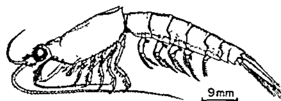


Fig. 2. *Lophogaster typicus*, adult male.

tween molts are shorter, maturity is attained earlier, at a smaller size, and life is shorter. In temperate zones, length averages 12–20 mm, and the life span averages 12–18 months. In the tropics, most species are probably annuals or less. Adults of 4 mm length are known. Cold-water forms, like Gnathophausians, may attain 200 mm. Their life span is unknown (Fig. 2).

After liberating the young, the females molt, mate, and immediately lay again. Fertilization is external, sperm being shed into the space between pairing couples or, rarely, into the brood sac. Brood may be liberated within a week in high temperatures or carried for months in low temperatures.

Ecology. Mysids are not combatant, relying on agility, camouflage, sheltering under rocks or weeds, or burrowing in the sea floor to escape enemies Pelagic species in upper waters are usually transparent and, except for the cornea, are colorless. Most littoral species have a chromatophore system whereby color changes are made to harmonize with the background. These are the result of hormones released to circulate in the blood in response to light stimuli on the eyes. No parasitic forms are known.

During daylight, most mysids remain feeding at, or near, the bottom, migrating at dusk towards upper levels where the young are liberated. Towards dawn, adults descend again while the juveniles usually remain nearer the surface until mature. Nightly vertical migrations through 400 m occur in some deepwater forms. Horizontal migrations are common; movement is landward at dusk and seaward before dawn.

Economic importance. Many species are gregarious, especially when breeding, and form valuable food for fish. Tidal waters of coastal lagoons near Calcutta, around Java and Japan are regularly sieved and the resulting catch, mainly mysids, are extensively used by natives as human food.

Systematics. Two suborders are usually recognized for this group. The Lophogastrida, with 7 genera and 37 species, are all marine, with the majority of species widely distributed in the abyssal realm. Some are mesopelagic, usually over great depths with localized distribution. Close affinities with fossil Mysidacea are seen in the following characters: imperfect fusion of the sixth and seventh abdominal somites, 7 pairs of gills, 7 pairs of brood-lamellae, and natatory pleopods which are biramous and alike in sexes, with only traces of sexual modification. No statocyst is present, feeding is by maxillary filter or raptorial, and there is practically no secondary sexual differentiation.

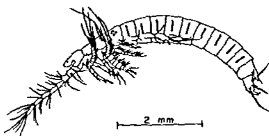
Fig 3 Mysida. (a) *Neomysis integer*, adult male. (b) *Mysidopsis gibbosa*, adult female.

Mysida, with 102 genera and 513 species, are widely distributed in all waters. A progressive specialization with loss of primitive characters is shown as follows; there is a complete fusion of the sixth and seventh abdominal somites in the adult. Gills are absent, and there is a reduction of the brood lamellae from 7 pairs in lower forms to 3 or 2 in higher. Female pleopods are rudimentary. They are progressively reduced in the male, with a transfer of natatory function to the thoracic exopods, and one or more pairs are sexually modified. Filter feeding is general, from ventral food currents, while raptorial feeding is supplementary. Males have a genital organ on the eighth thoracic appendage. Antennular brushes are present. A statocyst is present in all but the most primitive forms (Fig. 3). *See* MALACOSTRACA; PERACARIDA. [O.S.T.]

Bibliography: E. R. Lankester (ed.), *A Treatise on Zoology*, 1909; W. M. Tattersall and O. S. Tattersall, *The British Mysidacea*, Ray Society Monograph 136, 1951.

Mystacocarida

An order of primitive Crustacea, discovered in the interstices of Massachusetts intertidal sandy beaches in 1939. Specimens of both sexes have since been found in similar habitats on shores of Europe, Africa, and Chile. The three species are in the genus *Derocheilocaris*. The body is wormlike and 0.5 mm long. The cephalothorax bears first an-

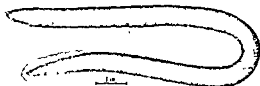


Mystacocarida, Derocheilocaris typicus. Lateral view of adult.

tennae, second antennae, mandibles, first maxillae, and second maxillae. Maxillipeds are on a separate segment and four additional free thoracic segments bear platelike appendages. The six abdominal segments are without appendages, but large caudal rami are present. The labrum is enormous, mouth parts and nervous system are primitive, and the genital pore is on the thorax. *See* CRUSTACEA. [R.W.PE.]

Myxiniformes

The hagfishes, Myxiniformes or Myxinoidea, constitute one of two orders of Recent jawless fishes. They are distantly related to the lampreys (Petromyzontiformes), which they resemble in the eel-like form. Hagfishes differ from lampreys in having



Atlantic hagfish, Myxine glutinosa. (After G. B. Goode, Great International Fisheries Exhibition, London, 1883, U S Natl. Museum Bull. 27, 1884)

a small mouth surrounded by enlarged fleshy barbels, the nostril opening to the outside below the upper lip and penetrating the palate internally, 5–15 pairs of gill pouches that open internally into the pharynx and externally through 1–15 pairs of pores; and in lacking visible eyes. The dorsal fin is continuous with the caudal fin and there is a single pair of semicircular canals. Hagfishes are hermaphroditic and oviparous. Development is direct and there is no larval stage.

[illegible]

flesh. When captured they secrete slime copiously, hence the vernacular name slime eels. *See* CYCLOSTOMATA (CHORDATA). [R.M.B.]

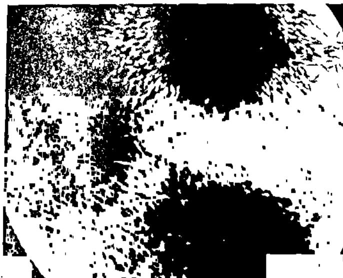
Myxobacterales

An order of bacteria of the class Schizomycetes. The Myxobacterales or myxobacters are also known as the slime bacteria. They are found mainly in the surface layers of soil, rotting wood, compost, and manure. *See* MANURE; SCHIZOMYCETES.

The vegetative cells are flexible, slender rods 1–2 μ wide and 3–5 μ long. The cells are not enclosed by a rigid cell wall; consequently, they are not as refractile as the cells of true bacteria and are somewhat difficult to see microscopically. On a surface the cells display a characteristic gliding motility, although neither flagella nor other motility organelles are present. Because of the motility of the cells, myxobacterial colonies are not limited like bacterial colonies but are, rather, rapidly spreading, thin "swarms," often with irregular, veinlike thickenings. The cells tend to arrange themselves in parallel rows, forming small tongues and islets at the edge of the swarm.

All members of the order, except in the genus *Cytophaga*, have a resting cell called the microcyst. The microcysts of most myxobacters are similar to the vegetative cell, being somewhat shorter and thicker than the latter. In the *Myxococcaceae*, however, the microcysts are spherical and are surrounded by a thick wall. Numerous microcysts may be contained in a larger structure called the cyst.





(a) Vegetative cells of *Myxococcus fulvus*. Note the typical parallel arrangement of the cells. (b) Edge of a colony of *Myxococcus fulvus* showing swarming of cells away from the colony. (*After N. A. Woods from A. T. Henrici and E. J. Ordal, The Biology of Bacteria, 3d ed., Heath, 1948*)

In fruiting myxobacters, the microcysts are found in structures called fruiting bodies.

Motility. The gliding motility shown by myxobacterial cells is found in several other groups of microorganisms, notably the blue-green algae and the Beggiatoales in the Schizomycetes. Myxobacters can glide only on a surface and are unable to swim in a liquid. The mechanism of this form of motion is unknown. The gliding cells deposit a layer of slime behind themselves. The direction of motion is generally parallel to lines of stress in the substrate; this response is called thigmotaxis. It is exhibited by several other organisms, including the bacterium *Kurthia zopfii* and mammalian fibroblasts. *See* BACTERIAL MOTILITY.

Microcysts and fruiting bodies. Microcysts of myxobacters are not comparable to bacterial endospores in heat resistance but can withstand considerable desiccation. *See* BACTERIAL ENDOSPORES.

The fruiting bodies formed by the myxobacters are unique among the bacteria. After a period of vegetative growth on a suitable medium, the cells of the swarm begin to aggregate toward several centers. The aggregating myxobacters form tongues of parallel rows of cells which converge on the center. Arriving at the center, the cells heap up on themselves and become transformed into microcysts. The fruiting body may be a more or less formless mass of microcysts embedded in a mass of slime or it may become variously branched and twisted, bearing clusters of cysts at the tips of the branches. Mature fruiting bodies are usually pigmented. They range in size from 0.1 to 1.0 mm or more. The only parallel to this remarkable life-cycle is found among the slime molds (Acrasiales), however, the vegetative development of the two groups is so different as to preclude any close relationship between them. *See* ACRASIALES.

Physiology and isolation. The physiology of the myxobacters has not been studied in any detail. Many of the *Cytophaga* and some of the fruiting forms can decompose cellulose and other complex polysaccharides. Except for some marine *Cytophaga*, all myxobacters are terrestial. *See* POLYSACCHARIDE.

Ignorance of the myxobacters' physiology renders their isolation highly empirical. The cellulose decomposing forms may be isolated and kept on a



Serial photographs showing a developing fruiting body of *Chondromyces crocatus*; the pictures were taken at hourly intervals. (*From J. T. Bonner, Morphogenesis, Princeton Univ. Press, 1952*)

simple medium containing the polysaccharide. For the majority of myxobacters perhaps the most successful method of isolation is to place several pieces of sterile rabbit dung on a layer of soil which is kept moist. After a few weeks, fruiting bodies will be visible on the dung and may then be transferred to a suitable medium.

Affinities of the myxobacters. The relationship of the myxobacters to other microorganisms is not clear. They resemble bacteria in size, in reproduction by binary fission, in their simple nucleus, and in the presence of the amino acid, diaminopimelic acid. In all these respects they also resemble the blue-green algae. They are set apart from the bacteria by their gliding motility, the lack of flagella, and the absence of a rigid cell wall. These properties recall the blue-green algae as well as the Beggiatoales among the Schizomycetes. On the other hand, the development of microcysts and especially of fruiting bodies is a unique feature of the myxobacters. Certainly, a very great deal more work must be done before the position of these organisms becomes clear. *See* BEGGIATOALES; CYANOPHYTA.

Classification. The order is divided into five families containing 13 genera. A brief description of each follows.

The Cytophagaceae do not have microcysts or other resting forms. They are often active in cellulose and chitin decomposition.

The Myxococcaceae have spherical microcysts, surrounded by a thick wall, and cysts may or may not be present. Usually simple fruiting bodies are formed except in the genus *Sporocytophaga*.

In the Archangiaceae, microcysts are not spherical, surrounded by a wall, or found in cysts. Fruiting bodies are formed.

In the Sorangiaceae, microcysts are not spherical or surrounded by a wall, but are contained in angular cysts. Fruiting bodies are formed.

In the Polyangiaceae, microcysts are not spherical or surrounded by a wall; they are contained in round or oval cysts. Usually, complex fruiting bodies are formed. [W.R.S.]

Myxomatosis, infectious (rabbits)

A rapidly fatal viral disease of European rabbits characterized by the formation of mucoid tumors in the skin and invasion of the blood and internal organs. It is also known as rabbit pox and is spread by the bite of the mosquito. Rabbits in the area where the disease originated are immune. It has been successfully used as a method of extermination of unwanted rabbits in Australia where laboratory-infected rabbits have been released into the rabbit population. It has been responsible for killing up to 100,000,000 rabbits, thus saving much grazing land.

The disease entity probably belongs to the pox family, and inclusion bodies can be found in the skin. *See* ANIMAL VIRUS; INCLUSION BODIES (VIRUS). [A.E.M.]

Myxomycetes

A group of microorganisms, also known as slime molds, comprising a class of the division Myxomycophyta of the plant kingdom. The organisms are on the borderline of the plant and animal kingdoms, with a noncellular, multinucleate, jellylike, creeping, assimilative phase, the plasmodium, which exhibits the structure and physiology of certain protozoa, and a fixed propagative stage of intricate, generally minute, often beautifully colored, fruiting bodies bearing walled spores. Because of the plasmodium, the slime molds are considered by some authorities to be Mycetozoa, an order of the phylum Protozoa. *See* ACRASIALES; LABYRINTHULALES.

The Myxomycetes are divided into the following subclasses, orders, and subfamilies:

Subclass Ceratiomyxetidae
  Order Ceratiomyxales
    Family Ceratiomyxaceae
Subclass Myxomycetidae
  Order Liceales
    Family Liceaceae
      Cribrariaceae
      Reticulariaceae
  Order Trichiales
    Family Dianemaceae
      Trichiaceae
  Order Stemonitales
    Family Collodermataceae
      Stemonitaceae
  Order Echinosteliales
    Family Echinosteliaceae
  Order Physarales
    Family Physaraceae
      Didymiaceae

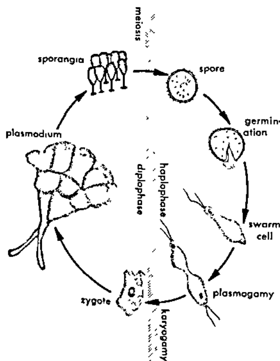
Myxomycetes abound in moist, shady places in all parts of the world, except possibly the more extreme polar regions. Most of the 400 or more known species are widely distributed, but some appear to be limited to tropical or temperate zones in their distribution. The plasmodia, which creep under the bark of rotting logs or in leaf mold on the forest floor, generally emerge just before fruiting, come to rest, and produce spore cases on the surface of logs, sticks, decaying leaves, fungus fruiting bodies, or even on the stalks of living plants. The stimuli which cause a plasmodium to sporulate are not clearly understood. Laboratory studies indicate that temperature, pH, moisture, food supply, and for some species, light all play major roles in inducing fruiting.

The life cycle of the Myxomycetidae, a subclass of the Myxomycetes, is fairly well understood. Under favorable conditions, the spores germinate, releasing 1–4 biflagellate swarm cells which fuse in pairs and form zygotes, or which lose their flagella and fuse as myxamebae. Nuclear fusion follows. Growth of the zygotes is accompanied by successive nuclear divisions resulting in the formation of

a plasmodium. An aggregation of zygotes appears to be a common occurrence in many species. Thus, a plasmodium may be formed from a single zygote by growth, or by the merging of several zygotes. After a period of development, the plasmodium changes into one or many fruiting bodies. A single plasmodium may form hundreds of sporangia each containing many spores.

There appear to be several distinct types of plasmodia in the Myxomycetes. The best known type, exemplified by *Physarum polycephalum* and related species, consists of a sheet of protoplasm with many-branched, veinlike strands which typically form a network with a fan-shaped margin. The veins consist of an outer layer surrounding a protoplasmic stream which exhibits a characteristic rhythm. The protoplasm streams in one direction for 50–90 seconds, stops momentarily, and then reverses the direction of flow for approximately the same length of time. G. W. Martin convincingly pointed out the fundamental similarity between the plasmodium of the Myxomycetes and the coenocytic mycelium of the Phycomycetes, and in his monograph of the North American Myxomycetes, the latter are treated as a class in his Division Fungi.

A typical fruiting body consists of a horny or lime-encrusted base (hypothallus), a stalk, and a spore case, the sporangium. The sporangial wall, the peridium, is noncellular. Inside the sporangium, the spores are enmeshed in a mass or network of threads, the capillitium, which in many species is elastic and, upon the rupture of the peridium, expands, elevating the spores. The presence or absence of any of these structures, and the characteristic form they take when present, are the criteria used in classifying the Myxomycetes. Species with brightly colored spores in which the sporangia are covered with a well-developed peridium are placed in the order Liceales if they have no capillitium, and in the Trichiales if capillitium is present. Species with dark spores are placed in the order Physarales if lime is present on the peridium or on the capillitium, but in the Stemonitales if no lime is present in either of these structures. Thus, the classification of the Myxomycetes is based chiefly on artificial characters, and no phylogenetic system is possible without further knowledge of structure, cytology, and physiology.



Life cycle of a myxomycete.

In their assimilative stages (myxamebae, plasmodia), Myxomycetes obtain their nutrition by engulfing and digesting bacteria, yeasts, and other microorganisms. In the laboratory, ordinary myxomycete cultures are fed rolled oats, but it is not certain whether the plasmodia obtain their nutriment directly from the oats, from the bacteria which attack the oat flakes, or from both.

It was not until 1939 that undoubted pure cultures of Myxomycetes were obtained. Even so, very few species have been grown in pure culture or even in two-membered culture with a known species of bacteria or yeasts. In fact, of the 400-odd known species, only 25–30 have been induced to complete their entire life cycle in artificial culture under any circumstances. Whether Myxomycetes in general are able to absorb and utilize food materials in solution is not known. *Physarum polycephalum*, which has been the subject of most physiological work, is the only myxomycete which has been grown in pure culture, in a chemically defined medium, in the absence of particulate food.

Because of their peculiar life cycle, the Myxomycetes are of great interest biologically and have become important experimental tools in the study of the structure of protoplasm and of the physiology of reproduction. [C.J.A.]

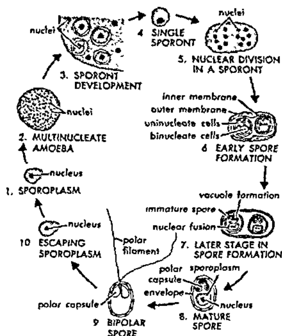
Bibliography: C. J. Alexopoulos, *Introductory Mycology*, 1952; A. L. Cohen, *Botan. Gaz.*, 101:243–275, 1939, 103:205–224, 1941; G. W. Martin, *Botan. Gaz.*, 93:421–435, 1932, *Botan. Rev.*, 6:356–388, 1940, *N. Am. Flora*, 1(1):1–190, 1949.

Myxosporidia

An order of Cnidosporidia which are characterized by the production of spores with one or more valves and polar capsules, and by possession of a single sporoplasm with or without an iodinophilous vacuole. Myxosporidians are mainly protozoan parasites of fishes. They infect all parts of the body, including the heart and brain, and often induce considerable pathological changes in the host tissue.

Most species have 2 valves and 2 polar capsules, and are further identified by the size and shape of the spore. The size, shape, and position of the capsules are of diagnostic value, as are also various

markings, prolongations, and appendages of the spore membrane. The exceptions to this general description are *Unicapsula*, with 2 valves and 1 polar capsule; *Chloromyxum*, with 2 valves and 4 polar capsules; *Kudoa*, with 4(?) valves and 4 polar capsules; and *Hexacapsula*, with 6 valves and 6 polar capsules.



Life cycle of Myxobolus.

These 4 genera contain species that cause extensive deterioration of the flesh in the Pacific halibut, the common swordfish, the North Atlantic herring and the Pacific yellowfin tuna, often making such commercially important fishes unfit for human consumption. Members of the genus *Myxobolus* may induce massive tumorlike growths in the skin of fresh- and brackish-water fishes. None, however, are pathogenic to humans or other mammals.

Infection begins with the ingestion of the spore by a host fish. The digestive fluids cause the polar filaments to be extruded, and at the same time the sporoplasm is released from the spore. The sporoplasm, or amebula, reaches the specific site of infection directly through the gut wall or by way of the blood stream. The amebula becomes a trophozoite when it starts feeding on the host tissues. The trophozoite then goes through a series of nuclear divisions (nucleogony) and, by a process of budding, gives rise to a number of cells, each of which eventually develops into a sporont. A sporont is a monosporoblast if one spore is produced and a pansporoblast if two or more spores are formed. The sporont undergoes a series of nuclear divisions, in which the number of nuclei produced will depend on the number of spores and polar capsules to be formed. Thus, for every spore a nucleus is involved in the formation of each valve and polar capsule. Two nuclei become the gametic nuclei which then fuse to form the zygotic nucleus of the sporoplasm. See CNIDOSPORIDIA. [R.F.N.]

Myxovirus

A term for a group of viruses that includes influenza, fowl plague, Newcastle disease, and mumps viruses. They are similar in size (about 100 mμ) and structure and resemble each other in possessing an enzyme capable of splitting off neuraminic acid derivatives from certain mucoproteins, a reaction concerned with the adsorption and penetration of virus into the cell. Myxoviruses agglutinate red blood cells, particularly those of man, guinea pig, and chicken. Such hemagglutination by the virus, and the development in the patient of specific antibodies which inhibit hemagglutination are properties useful in diagnosis. The myxoviruses are the only hemagglutinating viruses which, after absorption to red cells, spontaneously dissociate from them. This dissociation takes place because the virus enzyme destroys receptors at the cell surface. See ANIMAL VIRUS; INFLUENZA; MUMPS; NEWCASTLE DISEASE [J.L.M.]

Bibliography: T. M. Rivers and F. L. Horsfall, Jr. (eds.), *Viral and Rickettsial Infections of Man*, 3d ed., 1959.